

ISSUES IN THE ACCOMMODATION OF MODEL UNCERTAINTY IN  
MACRO-ECONOMETRIC MODELLING

**CHRYSTALLENI ARISTIDOU**

Thesis submitted to the University of Nottingham

For the Degree of Doctor of Philosophy

July 2016

Dedicated to my parents and husband

*For their endless love and unconditional sacrifice.*

# **Abstract**

This thesis deals with different types of uncertainty in various macroeconomic contexts and investigates ways in which these can be accommodated by adopting flexible techniques that allow a robust inference in estimation, testing and prediction. This thesis covers a wide range of aspects in macroeconomic analysis, including the choice of an appropriate unit root test, inference when the presence of breaks and the autocorrelation properties of data are unknown, characterisation of inflation dynamics when structural and specification uncertainty are present, as well as model uncertainty in forecasting when real-time data are available.

Chapter 1 presents the general motivations and describes the main research objectives and methodology for each chapter, providing a thesis outline at the same time.

Chapter 2 we examine the behaviour of OLS-demeaned/ detrended and GLS-demeaned/ detrended unit root tests that employ stationary covariates in situations where the magnitude of the initial condition of the time series under consideration may be non-negligible. We show that the asymptotic power of such tests is very sensitive to the initial condition; OLS- and GLS- based tests achieve relatively high power for large and small magnitudes of the initial condition, respectively. Combining information from both types of test via a simple union of rejections strategy is shown to effectively capture the higher power available across all initial condition magnitudes.

In Chapter 3, we consider a two-step procedure for estimating level break size(s) when the presence of the structural break(s) is uncertain and when the order of integration of the data is unknown. In other words, we deal with uncertainty over the appropriate filtering of the data, as well as structural uncertainty over the existence of a break. Our

approach is motivated by the well known interplay between the unit roots and structural changes: Evidence in favour of unit roots can be a manifestation of structural changes and vice versa. The proposed procedure is shown to exhibit substantial accuracy gains in estimating the level break-size and breakpoint.

Chapter 4 provides a characterisation of U.S. inflation dynamics within a generalised Phillips Curve framework that accommodates uncertainties about the duration a given Phillips Curve holds and the specification of the relationship, in addition to parameter and stochastic uncertainties accommodated within a typical Phillips Curve analysis. Our approach is based on an innovative method to deal with such uncertainties based on Bayesian model averaging techniques. Employing data for the U.S. in the period 1950q1-2012q4, the estimated version of the "meta" Phillips Curve provides an interesting characterisation of inflation dynamics which is in accordance with a number of distinguished studies.

Chapter 5 investigates the extent to which nowcast and forecast performance is enhanced by the use of real-time datasets that incorporate past data vintages and survey data on expectations in addition to the most recent data. The paper proposes a modelling framework and evaluation procedure which allow a real-time assessment and a final assessment of the use of revisions and survey data judged according to a variety of statistical and economic criteria. Both survey data and revisions data are found to be important in calculating density forecasts in forecasting the occurrence of business cycle events. Through a novel "fair bet" exercise, it is shown that models that incorporate survey and/or revisions data achieve higher profits in decision-making. The analysis also highlights the need to focus on future growth and inflation dynamics relevant to decision-makers rather than relying on simple point forecasts.

## Acknowledgements

I would like to express my deepest gratitude to my supervisors, Professor Kevin Lee and Steve Leybourne for their constant support and encouragement. Professor Kevin Lee has been a great inspiration and I owe him a deep debt of gratitude for his confidence in me and for sharing his immense experience. The innumerable suggestions and valuable comments from Professor Steve Leybourne substantially contributed to my studies. I sincerely thank him for his patience to guide me through and for always being there for me. I also want to express special thanks to Professor Robert Taylor, my supervisor in my first year, for his encouragement in my early steps as a PhD student. I also give thanks to Professor David Harvey and Professor Kalvinder Shields for their precious assistance.

I also would like to express my wholehearted thanks to my parents, Aristos and Despina, for their unconditional sacrifice and generous support throughout my entire life. Their encouragement gave me the strength to continue the struggle and accomplish my dreams. I owe profound gratitude to my husband, Dimitris, for his endless giving and understanding in the countless moments of crisis.

I also express my heartfelt gratefulness to my brother and sister-in-law for always putting a smile on my smile. Without their support, this thesis would not have been possible. I am also appreciative to my entire family and friends for their warm prayers and meaningful advice. Finally, I would like to express many thanks to Economic and Social Research Council for their scholarship and financial support throughout my studies.

# Table of Contents

Abstract .....	iii
Acknowledgements .....	iv
<b>1 Introduction .....</b>	<b>1</b>
1.1 Chapter 2: Initial Condition and Testing Choice Uncertainty .....	2
1.1.1 Motivation and Research Questions .....	2
1.1.2 Methodology .....	2
1.2 Chapter 3: Structural Breaks or Unit Roots? .....	3
1.2.1 Motivation and Research Questions .....	3
1.2.2 Methodology .....	4
1.3 Chapter 4: Inflation Dynamics and Structural Uncertainty .....	5
1.3.1 Motivation and Research Questions .....	5
1.3.2 Methodology .....	6
1.4 Chapter 5: Forecast Evaluation and the Usefulness of Real-Time Data .....	6
1.4.1 Motivation and Research Questions .....	6
1.4.2 Methodology .....	8
1.5 References .....	9
<b>2 The Impact of the Initial Condition on Covariance Augmented Unit Root Tests .....</b>	<b>12</b>
2.1 Introduction .....	12
2.2 The model and Covariance Augmented Unit Root Tests .....	13
2.2.1 OLS-based statistics .....	15
2.2.2 GLS-based statistics .....	16
2.3 Asymptotic Results .....	18
2.4 A Union of Rejections Strategy .....	30
2.5 Finite Sample Performance .....	35
2.6 Empirical Illustration .....	41
2.6.1 Unit Root Test on Industrial Production .....	42
2.6.2 Unit Root Test on CPI Inflation .....	47
2.7 Conclusion .....	51

2.8	Appendix A .....	51
2.9	References .....	59
3	<b>Dealing with Uncertainty over both the Existence of structural Breaks and Autocorrelation Properties of the Data .....</b>	<b>61</b>
3.1	Introduction .....	61
3.2	Literature Review .....	64
3.2.1	Estimation and Inference about Break Dates .....	64
3.2.2	Tests for Structural Change .....	68
3.3	One Level- Break .....	73
3.3.1	Model and Estimation .....	73
3.3.2	Finite Sample Performance .....	77
3.4	Two Level-Breaks .....	80
3.4.1	Model and Estimation .....	81
3.4.2	Finite Sample Performance .....	84
3.5	Empirical Illustration .....	85
3.6	Conclusion .....	87
3.7	Figures .....	89
3.8	References .....	118
4	<b>Regime Change, Inflation Dynamics and the meta-Phillips Curve .....</b>	<b>123</b>
4.1	Introduction .....	123
4.2	Literature Review .....	124
4.2.1	The New Keynesian Phillips Curve .....	124
4.2.2	The Hybrid New Keynesian Phillips Curve .....	126
4.2.3	Issues in Empirical Application .....	127
4.2.4	Regime Changes and the Phillips Curve .....	129
4.3	Modelling Framework .....	133
4.3.1	Bayesian Model Averaging .....	134
4.3.2	Estimation .....	138
4.4	Empirical Analysis .....	139
4.4.1	Data .....	139

4.4.2	Results .....	140
4.4.3	Robustness Check .....	153
4.5	Conclusion .....	155
4.6	Appendix A: Confidence Bands for Estimated Parameters in the meta-Phillips Curve .....	156
4.7	Appendix B: Larger Instrument Set .....	159
4.8	Appendix C: Smaller Instrument Set .....	161
4.9	Appendix D: Larger Probability of Break .....	164
4.10	References .....	166
<b>5</b>	<b>Nowcast and Forecast Evaluation when Real-Time Data are Available .....</b>	<b>172</b>
5.1	Introduction .....	172
5.2	Assessing the Usefulness of Real-Time Information .....	176
5.2.1	A Modelling Framework to Accommodate Real-Time Information .....	176
5.2.2	Assessing Nowcasting/ Forecasting Performance .....	179
5.3	Nowcasting Recessionary Events using US Real-Time Data .....	191
5.3.1	Real-Time Evaluation of Point and Density Forecasts .....	193
5.3.2	Final-Assessment of Model Predictability based on Economic Events .....	199
5.4	Structural Breaks and the Usefulness of Real-Time Data .....	207
5.4.1	Introducing Structural Breaks in the Real-Time Assessment .....	209
5.4.2	Results .....	214
5.5	Conclusion .....	225
5.6	Appendix A .....	226
5.7	References .....	230



# Chapter 1

## Introduction

Policymakers are continually faced with the eminently important issue of model uncertainty when designing economic plans or tailoring decisions, with different models often resulting in significantly different policy prescriptions. Structural breaks, flawed econometric practises, as well as fundamentally complex mechanisms embedded in economic environments, are among many factors complicating the model selection process. Relying on a single model, without paying the required attention to the uncertainty surrounding the true data generating process, can result in ambiguous outcomes that often entail severe macroeconomic repercussions. For this reason, research was directed towards accommodating model uncertainty in applied macro-econometric modelling, with the model averaging technique becoming increasingly attractive and empirically relevant. This thesis explores innovative model averaging techniques in various macroeconomic contexts, ranging from structural breaks inferences and inflation dynamics to forecasting using real-time data. At the same time, emphasis is placed in making valid inference in unit root testing, where uncertainty surrounding the true data generating process can also dramatically undermine the credibility of test outcomes.

Since this thesis covers a wide range of aspects in macroeconomic analysis, the chapters evolve around independent research questions, albeit linked with the common axiom of model uncertainty. In effect, the diversity of macro-econometric topics addressed in this thesis creates a stimulating framework in which we explore innovative and robust techniques that deal with uncertainty in various contexts which are often critical for empirical researchers. The remainder of this chapter provides the motivation, exact research questions and methodology for each of the main thesis chapters.

## 1.1 Chapter 2: Initial Condition and Testing Choice Uncertainty

### 1.1.1 Motivation and Research Questions

Research for this chapter has been motivated by the observation that different unit root tests behave differently depending of the structure of time series under consideration. Effectively, different unit root tests can result in different test outcomes, creating uncertainty regarding the true time-series properties of the data. In particular, Müller and Elliott (2003) show that tests based on the OLS-demeaning/detrending procedure of Dickey and Fuller (1979) and the GLS- demeaning/detrending procedure of Elliott, Rothenberg and Stock (1996) exhibit very different power profiles depending on the magnitude of the initial condition (the deviation of the initial observation from the underlying mean/trend in the data). In this chapter we examine the behaviour of covariate augmented unit root test over a range of initial condition values and propose a robust procedure for mitigating against the uncertainty arising from different test outcomes.

Hansen (1995) has shown that augmenting the traditional OLS regression model with that covariate can substantially increase the power of the OLS-based unit root test while Elliott and Jansson (2003) and Westerlund (2013) demonstrate that similar advantages are obtained in a GLS-demeaning/detrending setting. In this chapter we demonstrate that for a small initial condition, the GLS-based test (Elliott and Jansson (2003) test) can have substantially more power than its OLS- based counterpart (Hansen (1995) test), while the reverse is true for a large initial condition. Typically, the power of the OLS-based test is an increasing function of this magnitude, whereas the GLS-based test demonstrates the opposite behaviour.

The initial condition is not known and it cannot be consistently estimated. It cannot be said whether the initial condition is small or large since different conclusions are reached depending on the sample window considered. This creates considerable uncertainty regarding the choice of the appropriate unit root test as it is not clear which of the two types of test to apply.

### 1.1.2 Methodology

To accommodate uncertainty regarding the choice of the covariate augmented unit root test, we follow Harvey et al. (2009a) who propose a simple union of rejections strategy whereby the unit

root null hypothesis is rejected whenever either of the individual OLS- or GLS- based unit root tests rejects. This methodology is shown to capture the superior power properties of the GLS-based test for a small initial condition and the superior power properties of the OLS-based test for a large initial condition. Accordingly, a union of rejections decision rule between the Elliott and Jansson's test and Hansen's test is fruitfully explored in order to maintain good power properties across both large and small initial conditions. Reported asymptotic evidence suggests that our approach offers good robust power performance in the presence of uncertainty over the magnitude of the initial condition, retaining attractive power levels across zero, small and large initial condition magnitudes.

## **1.2 Chapter 3: Structural Breaks or Unit Roots?**

### **1.2.1 Motivation and Research Questions**

Structural breaks uncertainty can arise when the break size is small and difficult to detect or locate. If there is a break and the researcher fails to account for it, then this can have detrimental effects on modelling, estimation and forecasting. The usual practice in applied econometrics is to perform a structural break test, such as the test proposed by Andrews (1993) and Andrews and Ploberger (1994). If the test rejects the null of no structural break, then estimation and inference is carried out conditional on the presence of the break. However, Hansen (2009) highlights the poor sampling properties associated with pretesting routines while Harvey, Leybourne and Taylor (2012) show that unit-root tests which are based on auxiliary statistics for the detection of breaks can deliver power function with undesirable properties.

Concurrently, the detection and location of breaks can be obscured by uncertainty regarding the autocorrelation properties of the series, and in particular, whether the underlying process accepts a unit root or not. Research has been directed towards making structural breaks inferences when the order of integration is unknown (inter alia Vogelsang (1998), Harvey, Leybourne and Taylor (2009b, 2010) and Perron and Yabu (2009)). It has been suggested that one way to deal with this uncertainty is to pretest for a unit root and then conduct a structural breaks analysis by choosing the break fraction that minimises the sum of squared residuals from a regression on the level or the first-difference of the series, depending on the unit root outcome in the first phase. Carrion-i-

Silvestre, Kim and Perron (2009) also suggest choosing the break location based on quasi-differenced regressions. However, a unit root test will only inform the researcher whether a unit root is present and will not indicate the appropriate filtering of the data needed to construct the quasi-differenced series. In addition, a unit root test that does not account for structural breaks when they occur can substantially distort inferences since the power of such tests is severely reduced when the process is stationary but subject to breaks (Perron (1989)). Effectively, whether the emphasis is placed on unit roots or structural breaks, the interplay between structural breaks and unit root tests poses a threat to valid inference.

Consequently, the motivation for this chapter is the interplay between unit root tests and structural breaks: Evidence in favour of unit roots can be a manifestation of structural breaks and vice versa. Should the researcher perform a unit-root test before testing for the presence of structural breaks or should he/she test for the presence of structural breaks before conducting unit root tests? Accordingly, this chapter considers break size estimation when the presence of breaks is uncertain and the order of integration of the data is unknown.

### 1.2.2 Methodology

We propose a two-step procedure that simultaneously deals with uncertainty over the existence of breaks and the autocorrelation properties of data. In the first step, we follow Harvey and Leybourne (2013) whose procedure selects between first differences estimators and a number of quasi-differenced alternatives according to which achieves the smallest minimum sum of squared residuals, mitigating against the possibility of erroneously selecting purely between the level of the series and first-differenced series. The method is particularly conservative since a break is included a priori, so that the choice of the filtering parameter is considered credible, even when no break has actually taken place.

The second step exploits the advantages of model averaging in estimation, so that the most robust inferences can be obtained when the existence of break is uncertain. After applying the selected filtering, as indicated in the first step, we distinguish between a restricted and unrestricted model. The restricted model specifies that no break takes place, so that the break size is zero, while the unrestricted model allows for a break. The second step involves taking a weighted average of

the break size under the two models with the weights selected to minimise a modified Mallows information criterion (Hansen (2009)). It is found that the approach attains substantial accuracy gains relating to the location of the break and its size. This superiority over Hansen’s (2009) original proposal is a result of the application of an appropriate filtering parameter.

## **1.3 Chapter 4: Inflation Dynamics and Structural Uncertainty**

### **1.3.1 Motivation and Research Questions**

This chapter considers structural instability of the new-Keynesian Phillips Curve and examines the suitability of various driving force measures that can enter the relationship. The aim of this chapter is to characterise the U.S. inflation dynamics accommodating structural instability arising from regime breaks and changes in the underlying drivers of price-setting decisions.

The potential of structural instability, which arises as central banks revise their monetary policies according to the macroeconomic environment that prevails, creates significant uncertainty regarding the response of inflation to expectations, past inflation and the real activity. The chapter is motivated by the numerous and dramatic changes that inflation dynamics have undergone through: The abrupt shifts in the level of inflation, the flattening of the Phillips Curve and the decline in inflation persistence are likely to be important determinants of the changing inflation dynamics, establishing different inflation regimes. Accordingly, we expect that this “regime uncertainty” makes the choice of the appropriate sample window a challenging task.

Moreover, work on this chapter has been motivated by the ongoing debate about which variable that captures changes in the real economy should enter the Phillips Curve relationship. The real marginal cost that enters the relationship is an unobservable variable that has to be estimated. Several authors, such as Fuhrer and Moore (1995a, b) and Neiss and Nelson (2002), suggest using the output gap as a measure of capacity utilisation but the use of the output gap does not deliver theoretically consistent results as emphasised by Galí and Gertler (1999). Instead, Sbordone (2002) Galí et al.(2001) and Galí and Gertler (1999) suggest using labour share, documenting that this measure incorporates both productivity and wage pressures to influence inflation. Here, the gap created by this controversy is investigated in a flexible way, using model averaging techniques.

### **1.3.2 Methodology**

Against this backdrop, this chapter provides an interesting characterisation of U.S. inflation dynamics that embeds the substantial uncertainties arising from structural breaks and specification issues, through robust model averaging techniques, proposed by Lee et al. (2015). The approach involves a set of specific Phillips Curves which are distinguished by the sample size over which they are estimated as well as the specific measure of real marginal cost that enters the relationship. The idea is that if there is a break, a new Phillips Curve prevails, so that the sample period over which it holds is small. If there is regime continuity so that no break occurs, the sample size over which a given Phillips Curve holds is augmented by one additional observation as we move through the sample. The individual Phillips Curves are estimated through Generalised Methods of Moments but are combined using Bayesian Model Averaging techniques, employed in a recursive setup.

The model weights are determined by the ability of individual Phillips Curves to explain inflation dynamics, allowing them to evolve and providing the most reliable inferences of inflation dynamics in a flexible setting. In particular, a model's weight depends on the probability of observing the last observation in the sample and the probability that the model remains relevant, which itself depends on last period's weights and the transition probability which reflects the possibility of a break. The approach is shown to perform well in characterising inflation dynamics, capturing all the major changes in inflation history, while it provides important evidence on the usefulness of different measures of the real activity.

## **1.4 Chapter 5: Forecast Evaluation and the Usefulness of Real-Time Data**

### **1.4.1 Motivation and Research Questions**

The purpose of this chapter is to judge the usefulness of real-time data in decision-making, focusing on the use of revision and survey data in nowcasting and forecasting quarterly inflation and output growth in the U.S. using novel evaluation criteria and model averaging techniques. The advantages of model averaging in forecasting have been widely examined (see for instance Harvey and Newbold

(2005)) so one of the scopes of this chapter is to investigate these benefits in the real-time data context. Since first release data are published with a delay, agents often use information from surveys as direct measures of expectations of the contemporaneous and future values of the variables of interest. In addition, statistical agencies usually provide preliminary estimates of the variables at the earliest opportunity, subsequently revising them as more complete information flows in.

While the incorporation of survey and revisions data in real-time analysis has been widely examined, evidence on the usefulness of such data in forecasting has been mixed. Koenig, Dolmas and Piger (2003) argue that because researchers tend to mix heavily revised data with first-release and lightly revised data, the conventional approach of using all types of data is unlikely to yield good forecasts. On the other hand, Garratt et al. (2008), Clements and Galvão (2013) and Jacobs and van Norden (2011) all provide evidence that the forecast performance is significantly improved when modelling data revisions. Consensus has not been reached on the usefulness of survey data either. Among others, Matheson et al. (2010) and Bańbura and Rünstler (2011) demonstrate that expectations data enhance the predictive ability of models while Croushore (2010) reports evidence from studies that uncover the biases relating to expectational errors. This chapter addresses the mixed evidence in literature and sets out a modelling framework that accommodates revisions data and survey data alongside the first-release data.

In addition, research on this chapter is also motivated by the vast literature on forecast evaluation. Recently researchers have shifted attention from point forecasts to density forecasts in an attempt to incorporate forecast uncertainty, relying more on logarithmic scores than root mean squared errors when evaluating a model's forecast performance. Even more importantly, a strand of literature focuses on the economic value of a model's value, judging the performance of a model in specific decision making settings (see Granger and Machina (2006), Granger and Pesaran (2000)). Accordingly, we judge the usefulness of real-time data not just in terms of statistical evaluation criteria, but also their contribution in accumulating profits to investors predicting relatively rare events.

### 1.4.2 Methodology

To address this chapter's research aims, we adopt a simple modelling framework in which we incorporate revisions and survey data alongside the most recent data measures to characterise the data generating process of the variables of interest. The usefulness of revisions and survey data is judged according to the nowcasting and forecasting performance of models that fully make use of the data compared to the performance of models that only make partial use of information or that are based only on the most recent vintage of data. In doing so, we employ model averaging techniques in a recursive set-up, diversifying against structural breaks.

The analysis provides a real-time assessment and a final assessment of forecast performance. The real-time assessment involves a recursive estimation of individual VAR models, distinguished by the use of revisions and survey data and combined using model averaging methods. The weights reflect the relative nowcasting performance of each model according to statistical criteria such as the mean root squared error or logarithmic score. This real assessment exercise results in four average models: (i) the average model that uses both revisions and survey data to different extent, (ii) the average model that makes no use of revisions data, (iii) the average model that makes no use of survey data and (iv) the average model that makes no use of revisions and survey data.

The final assessment then judges the performance of the four average models based on statistical criteria, as well as economic criteria which are based on explicit investment decision making strategies. In the latter case, we focus on a set of events involving output and inflation that are frequently discussed in business cycle analysis and introduce a novel means of evaluating probability forecasts which is built on fair-bets.

This methodology is shown to deliver interesting research outcomes: Judged by statistical criteria, the performance of models that nowcast and forecast output growth and inflation is considerably enhanced by taking into account information available in survey data and revisions data. We also show that relating to economic significance, both revisions data and survey expectations are important in calculating density forecasts in forecasting the occurrence of a number of events. Moreover, the use of revisions and survey data can substantially boost investors' profits in decision-making based on forecasts, as shown by a "fair bet" exercise. In addition, the analysis demonstrates that



evaluating forecasts based on future growth and inflation dynamics that are relevant to decision-makers is a much more reliable practice than relying solely on point forecasts.

## 1.5 References

1. Andrews, D.W.K. (1993), "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, pp.821-856.
2. Andrews, D.W.K. and W. Ploberger (1994), "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62 (6), pp.1383-1414.
3. Carrion-i-Silvestre, J.L., D. Kim, and P. Perron (2009), "GLS-based unit root tests with multiple structural breaks both under the null and the alternative hypotheses," *Econometric Theory*, 25, pp.1754-1792.
4. Croushore, D. (2010), "An Evaluation of Inflation Forecasts from Surveys using Real-Time Data," *The BE Journal of Macroeconomics: Contributions*, 10(1), pp.1-32.
5. Dickey, D.A. and W.A. Fuller (1979), "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74, pp.427-431.
6. Elliott, G. and M. Jansson (2003), "Testing for unit roots with stationary covariates," *Journal of Econometrics*, 115, pp.75-89.
7. Elliott, G., T. J. Rothenberg and J. H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64 (4), pp.813-836.
8. Fuhrer, J. and G. Moore (1995a), "Monetary policy trade-off and the correlation between nominal interest rates and real output," *American Economic Review*, 85(1), pp.219-239.
9. Fuhrer, J. and G. Moore (1995b), "Inflation Persistence," *Quarterly Journal of Economics*, 110, pp.127-159.
10. Galí, J. and M. Gertler (1999), "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, 44, pp.195-222.

11. Galí, J., M. Gertler and D. Lopez-Salido (2001), "European inflation dynamics," *European Economic Review*, 45 (7), pp.1237-1270.
12. Garratt, A., K. Lee, E. Mise and K. Shields (2008), "Real time representations of the output gap," *Review of Economics and Statistics*, 90(4), pp.792-804.
13. Granger, C.W.J and M. Machina (2006), "Forecasting and Decision Theory," in G. Elliott, C.W.J. Granger and A.Timmermann (eds.), *Handbook of Economic Forecasting*, pp.81-98.
14. Granger, C.W.J. and M.H. Pesaran (2000), "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, 19, pp.537-560.
15. Hansen, B. E. (1995), "Rethinking the univariate approach to unit root testing using covariates to increase power," *Econometric Theory*, 11, pp.1148-1171.
16. Hansen, B. E. (2009), "Averaging estimators for regressions with a possible structural break," *Econometric Theory*, 35, pp.1498-1514.
17. Harvey, D. I. and S.J. Leybourne (2013), "Break Date Estimation for Models with Deterministic Structural Change," *Oxford Bulletin of Economics and Statistics*, 0305-9049, pp.1-20.
18. Harvey, D. I., S. J. Leybourne and A.M.R. Taylor (2010), "Robust methods for detecting multiple level breaks in autocorrelated times series," *Journal of Econometrics*, Vol.157, pp.342-358.
19. Harvey, D.I, S.J. Leybourne and A.M.R. Taylor (2012), "Unit root testing under a local break in trend," *Journal of Econometrics*, 167(1), pp. 140-167.
20. Harvey, D.I, S.J. Leybourne and A.M.R. Taylor (2009a), "Unit root testing in practice: Dealing with uncertainty over the trend and initial condition," *Econometric Theory*, 25, pp.587-636.
21. Harvey, D.I, S.J. Leybourne and A.M.R. Taylor (2009b), "Simple, robust and powerful tests of the breaking trend hypothesis", *Econometric Theory*, Vol.25, pp.995-1029.

22. Harvey, D.I. and P. Newbold (2005), "Forecast Encompassing and Parameter Estimation," Oxford Bulletin of Economics and Statistics (Supplement), 67, pp.815-835.
23. Jacobs, J. and S. van Norden (2011), "Modeling Data Revisions: Measurement Error and Dynamics of True Values," Journal of Econometrics, 161, 2, pp.101-109.
24. Koenig, E. F., S. Dolmas and J. Piger (2003) "The use and abuse of real-time data in economic forecasting," The Review of Economics and Statistics, 85(3), pp.618-628.
25. Lee, K., Morley J. and Shields K. (2015), "The meta Taylor Rule," Journal of Money, Credit and Banking, Vol.47, Issue 1, pp.73-98.
26. Matheson, T.D., J. Mitchell and B. Silverstone (2010), "Nowcasting and predicting data revisions using panel survey data," Journal of Forecasting, 29(3), pp.313-330.
27. Müller, U.K. and G. Elliott (2003), "Tests for unit roots and the initial condition," Econometrica, 71, pp.1269-1286.
28. Perron, P. (1989), "The great crash, the oil price shock and the unit root hypothesis," Econometrica 57, pp.1361-1401.
29. Perron, P. and T. Yabu (2009), "Testing for shifts in trend with an integrated or stationary noise component," Journal of Business and Economic Statistics, Vol.27, pp.369-396.
30. Sbordone, A. (2002), "Prices and Unit Labor Costs: A New Test of Price Stickiness," Journal of Monetary Economics, 49(2), pp.265-292.
31. Vogelsang, T.J. (1998), "Testing for a shift in mean without having to estimate serial-correlation parameters", Journal of Business and Economic Statistics, Vol.16, pp.73-80.
32. Westerlund, J. (2013), "A computationally convenient unit root test with covariates, conditional heteroskedasticity and efficient detrending," Journal of Time Series Analysis, 34, pp.477-495.

# Chapter 2

## The Impact of the Initial Condition on Covariate Augmented Unit Root Tests

### 2.1 Introduction

Conventional testing for a unit root in a time series is typically carried out using the OLS-demeaning/detrending procedure of Dickey and Fuller (1979), or the GLS-demeaning/detrending procedure of Elliott, Rothenberg and Stock (1996). When the series under consideration covaries with an available stationary variable, Hansen (1995) showed that it is possible to substantially increase the power of the OLS-based unit root tests by augmenting the underlying OLS regression model with that covariate. Elliott and Jansson (2003) and Westerlund (2013) show that incorporating covariates in a GLS-demeaning/detrending setting also improves the power of GLS-based unit root tests.

As shown in Müller and Elliott (2003), the powers of conventional OLS-based and GLS-based unit root tests are sensitive to the magnitude of the unobserved initial condition of a time series. For a small initial condition, GLS-based tests can have substantially more power than their OLS-based counterparts, while the reverse is true for a large initial condition. Typically, the power of OLS-based tests is an increasing function of this magnitude, whereas GLS-based tests demonstrate the opposite behaviour. In any practical testing situation, the magnitude of the initial condition is not known (nor can it be consistently estimated) and it is therefore unclear whether it is best to apply an OLS- or GLS-based unit root test in order to extract the most information about the

presence, or otherwise, of a unit root. Harvey et al. (2009a) examine the behaviour of a simple union of rejections strategy whereby (in its simplest guise) the unit root null hypothesis is rejected whenever either of the individual OLS- or GLS- based unit root tests rejects. This procedure is shown to perform well in practice since it captures the superior power of the GLS-based test for a small initial condition and the superior power of the OLS-based test for a large initial condition.

In this chapter, we show that the patterns of sensitivity of the power of OLS- and GLS-based covariate augmented unit root tests to the magnitude of the initial condition are actually very similar to that of their non-covariate augmented counterparts. This implies that the same considerations are relevant as in the non-covariate augmented case, when deciding which of the OLS- or GLS-based covariate augmented unit root tests to apply. Our proposed solution is once again to employ a union of rejections strategy, which we demonstrate is very effective in the covariate augmented context. The plan of the chapter is as follows. The next section sets out the model and describes the Hansen (1995) and Elliott and Jansson (2003) covariate augmented unit root tests. Here we also consider a simpler variant of the Elliott and Jansson (2003) GLS-based test following Westerlund (2013) which proves useful in the context of the union of rejections strategy. Section 2.3 derives the local asymptotic power functions of the tests in the presence of possibly non-negligible initial conditions and examines their asymptotic local powers. Section 2.4 introduces the union of rejections strategies and examines their large sample power properties. Finite sample power comparisons are shown in section 2.5, which also includes discussion of issues regarding the practical implementation of the recommended procedure. Section 2.6 includes an empirical illustration of the proposed strategy. Section 2.7 concludes the chapter. In what follows,  $\|(\cdot)$  denotes the indicator function,  $L$  denotes the lag operator,  $\xrightarrow{p}$  denotes convergence in probability and  $\Rightarrow$  denotes weak convergence.

## 2.2 The model and covariate augmented unit root tests

For purposes of transparency we will conduct our analysis within the context of a fairly simple model that admits a single covariate and abstracts from serial correlation in the innovations. We

consider the following model for the series  $y_t$  and the covariate  $x_t$ ,  $t = 1, \dots, T$ :

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mu_y + \beta_y t \\ \mu_x + \beta_x t \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ u_{x,t} \end{bmatrix} \quad (2.1)$$

where

$$\begin{bmatrix} u_{y,t} - \rho u_{y,t-1} \\ u_{x,t} \end{bmatrix} = \begin{bmatrix} v_t \\ e_t \end{bmatrix}. \quad (2.2)$$

Within this generic data generating process (DGP) specification we identify three alternative specifications for the deterministic components of  $y_t$  and  $x_t$ , with varying restrictions concerning the trend component of  $y_t$  and  $x_t$ :

$$\text{Model A} : \quad \beta_y = \beta_x = 0 \quad (2.3)$$

$$\text{Model B} : \quad \beta_y \neq 0, \beta_x = 0 \quad (2.4)$$

$$\text{Model C} : \quad \beta_y \neq 0, \beta_x \neq 0 \quad (2.5)$$

In Model A, no trends are assumed present in either  $y_t$  or  $x_t$ ; in Model B, a trend is permitted in  $y_t$  alone, while both  $y_t$  and  $x_t$  admit a trend in Model C. We make the following assumption regarding the innovations  $v_t$  and  $e_t$ :

**Assumption 1.** *The stochastic process  $\varepsilon_t = \begin{bmatrix} v_t & e_t \end{bmatrix}'$  is a martingale difference sequence with variance  $E(\varepsilon_t \varepsilon_t') = \Omega$  where*

$$\Omega = \begin{bmatrix} \sigma_v^2 & \sigma_{ev} \\ \sigma_{ev} & \sigma_e^2 \end{bmatrix} \quad (2.6)$$

and  $\sup_t E(\|e_t\|^4) < \infty$ . Let the squared correlation between the innovation  $v_t$  and the covariance  $u_{x,t} = e_t$  be denoted by

$$R^2 = \frac{\sigma_{ev}^2}{\sigma_v^2 \sigma_e^2}. \quad (2.7)$$

Within (2.2), for the autoregressive process  $u_{y,t}$  we set  $\rho = 1 + c/T$  for  $c \leq 0$ , with  $c = 0$  and  $c < 0$  corresponding to unit root and local-to-unit root autoregressive processes, respectively. Here  $u_{x,t} = e_t$  is the covariate which is correlated with the innovation term of  $u_{y,t}$  when  $\sigma_{ev} \neq 0$  (i.e. when  $R^2 > 0$ ).

In this chapter we wish to allow for the possibility that the initial condition of the autoregressive process  $u_{y,t}$ , i.e.  $u_{y,1}$ , is asymptotically non-negligible, so that its limiting effect on covariate aug-

mented unit root tests can be ascertained. Specifically, the following assumption is made regarding the behaviour of  $u_{y,1}$ :

**Assumption 2.** *For  $c < 0$ , the initial condition is generated according to  $u_{y,1} = \alpha \sqrt{\sigma_v^2 / (1 - \rho^2)}$ , where  $\alpha$  is a fixed parameter. For  $c = 0$ , we may set  $u_{y,1} = 0$  without loss of generality, due to the exact similarity of the covariate augmented unit root tests to the initial condition in this case.*

In Assumption 2, the parameter  $\alpha$  controls the magnitude of the fixed initial condition  $u_{y,1}$  (i.e. the deviation of the initial observation from the underlying mean/trend in the data) relative to the standard deviation of a stationary  $AR(1)$  process with parameter  $\rho$  and innovation variance  $\sigma_v^2$ . This form for the initial condition is closely related to that given in Müller and Elliott (2003) and Harvey and Leybourne (2005). Notice also that, when  $c < 0$ , the initial value is not asymptotically negligible because  $T^{-1/2}u_{y,1} \rightarrow \alpha\sigma_v/\sqrt{-2c}$  as  $T \rightarrow \infty$ .

Our focus in this chapter is on testing the unit root null  $H_0 : \rho = 1$  against the stationary alternative  $H_0 : \rho < 1$ , in the case where a covariate is available. In the context of the model (2.1)-(2.2) and Assumption 1 we now outline statistics that derive from the Hansen (1995) and Elliott and Jansson (2003) approaches to covariate augmented unit root testing, which are respectively based on OLS and GLS detrending of the  $y_t$  data.

### 2.2.1 OLS-based statistics

The Hansen (1995) approach tests for a unit root in  $y_t$  using a Dickey-Fuller-type regression, augmented by the covariate as an additional regressor, and implicitly employs OLS demeaning/detrending of the  $y_t$  and  $x_t$  series (note that Hansen does not consider Model C, but extension to this case is trivial). Based on our components representation of the DGP in (2.1), we express this type of statistic as follows:

$$t_{\hat{\phi}} = \frac{\hat{\phi}}{s.e.(\hat{\phi})} \quad (2.8)$$

where  $\hat{\phi}$  and  $s.e.(\hat{\phi})$  are the OLS estimate and associated standard error of  $\phi$  obtained from the regression

$$\Delta \hat{u}_{y,t} = \phi \hat{u}_{y,t-1} + \delta \hat{u}_{x,t} + \eta_t \quad (2.9)$$

with  $\hat{u}_{y,t}$  and  $\hat{u}_{x,t}$  denoting residuals from the OLS demeaned/detrended  $y_t$  and  $x_t$  series

$$\hat{u}_{y,t} = \begin{cases} y_t - \hat{\mu}_y & \text{for Model A} \\ y_t - \hat{\mu}_y - \hat{\beta}_y t & \text{for Models B, C} \end{cases} \quad (2.10)$$

$$\hat{u}_{x,t} = \begin{cases} x_t - \hat{\mu}_x & \text{for Models A,B} \\ x_t - \hat{\mu}_x - \hat{\beta}_x t & \text{for Model C} \end{cases} \quad (2.11)$$

where in the demeaned cases,  $\hat{\mu}_y$  and  $\hat{\mu}_x$  denote the estimated intercepts in the regressions of  $y_t$  and  $x_t$ , respectively, on a constant, while in the detrended cases,  $\hat{\mu}_y$ ,  $\hat{\beta}_y$  and  $\hat{\mu}_x$ ,  $\hat{\beta}_x$  denote the intercept, trend coefficient estimates in the regressions of  $y_t$  and  $x_t$ , respectively, on a constant and linear trend.

### 2.2.2 GLS-based statistics

Elliott and Jansson (2003) propose an approach to covariate augmented unit root testing based on a likelihood ratio principle combined with GLS demeaning/detrending for  $y_t$  but retaining OLS demeaning/detrending for the covariate  $x_t$ . Specifically, for our basic model, their statistic is given by:

$$\hat{\Lambda} = T \left\{ tr \left( \left[ \sum_{t=1}^T \hat{u}_t(1) \hat{u}_t(1)' \right]^{-1} \left[ \sum_{t=1}^T \hat{u}_t(\bar{\rho}) \hat{u}_t(\bar{\rho})' \right] \right) - 1 - \bar{\rho} \right\} \quad (2.12)$$

where, for  $r = \bar{\rho} = 1 + \bar{c}/T$  (for some chosen  $\bar{c} < 0$ ) and  $r = 1$ ,

$$\hat{u}_t(r) = z_t(r) - d_t(r)' \hat{\beta}(r) \quad (2.13)$$

with

$$z_t(r) = \begin{bmatrix} (1 - r \parallel (t > 1)L)y_t \\ x_t \end{bmatrix} \quad (2.14)$$

$$d_t(r)' = \begin{cases} \begin{bmatrix} 1 - r \parallel (t > 1) & 0 \\ 0 & 1 \end{bmatrix} & \text{for Model A} \\ \begin{bmatrix} 1 - r \parallel (t > 1) & 0 & (1 - r \parallel (t > 1)L)t \\ 0 & 1 & 0 \end{bmatrix} & \text{for Model B} \\ \begin{bmatrix} 1 - r \parallel (t > 1) & 0 & (1 - r \parallel (t > 1)L)t & 0 \\ 0 & 1 & 0 & t \end{bmatrix} & \text{for Model C} \end{cases} \quad (2.15)$$



and

$$\hat{\beta}(r) = \left[ \sum_{t=1}^T d_t(r) \hat{\Omega}^{-1} d_t(r)' \right]^{-1} \left[ \sum_{t=1}^T d_t(r) \hat{\Omega}^{-1} z_t(r) \right] \quad (2.16)$$

where  $\hat{\Omega}$  is a consistent estimator of  $\Omega$ .

An alternative approach to covariate augmented unit root testing that also makes use of GLS demeaning/detrending for  $y_t$  is to adapt the Hansen (1995) Dickey-Fuller-based statistic, where the deterministic coefficients in (2.1) are estimated using GLS rather than OLS, an approach suggested by Westerlund (2013). Specifically, we consider the following GLS-based variant of Hansen's statistic:

$$t_{\tilde{\phi}} = \frac{\tilde{\phi}}{s.e.(\tilde{\phi})} \quad (2.17)$$

where  $\tilde{\phi}$  and  $s.e.(\tilde{\phi})$  are obtained from the fitted OLS regression

$$\Delta \tilde{u}_{y,t} = \tilde{\phi} \tilde{u}_{y,t-1} + \tilde{\delta} \hat{u}_{x,t} + \tilde{\eta}_t \quad (2.18)$$

with  $\hat{u}_{x,t}$  denoting residuals from the OLS demeaned/detrended  $x_t$  series as before, but now  $\tilde{u}_{y,t}$  denoting the GLS demeaned/detrended  $y_t$  series, obtained from an OLS regression of  $(1 - \bar{\rho} \parallel (t > 1)L)y_t$  on  $1 - \bar{\rho} \parallel (t > 1)$  for Model A, and  $(1 - \bar{\rho} \parallel (t > 1)L)y_t$  on  $[1 - \bar{\rho} \parallel (t > 1), (1 - \bar{\rho} \parallel (t > 1)L)t]'$  for Models B and C.

Both the  $\hat{\Lambda}$  and  $t_{\tilde{\phi}}$  GLS-based statistics rely on specifying a value of  $\bar{c}$ . Elliott and Jansson (2003) and Westerlund (2013) suggest using the Elliott et al. (1996) values of  $\bar{c} = -7$  for Model A and  $\bar{c} = -13.5$  for Models B and C. These choices are motivated by the value of  $c = \bar{c}$  for which the nominal 0.05-level asymptotic Gaussian local power envelope is at 0.50 in the non-covariate augmented case, which corresponds to the case of  $R^2 = 0$  in the context of the covariate augmented tests. As Elliott and Jansson (2003) and Westerlund (2013) note, it is also possible to select  $\bar{c}$  according to the value of  $R^2$ , so that the asymptotic Gaussian local power envelope is at 0.50 for any given  $R^2$ , but these authors do not recommend such an approach, arguing that unit root test power is increasing in  $R^2$  (for a given  $c$ ), and so base their choice of a single  $\bar{c}$  parameter on the lowest power scenario ( $R^2 = 0$ ). In what follows, we follow such previous work and set  $\bar{c} = -7$  for Model A and  $\bar{c} = -13.5$  for Models B and C.

## 2.3 Asymptotic Results

In this section we derive the local asymptotic distributions for  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  under Assumptions 1 and 2, when  $\rho = 1 + c/T$ ,  $c \leq 0$ . We make use of the following weak convergence result

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \begin{bmatrix} v_t \\ e_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_v & 0 \\ \sigma_e R & \sqrt{\sigma_e^2(1-R^2)} \end{bmatrix} \begin{bmatrix} W_1(r) \\ W_2(r) \end{bmatrix} \quad (2.19)$$

$$= \begin{bmatrix} \sigma_v W_1(r) \\ \sigma_e \{RW_1(r) + \sqrt{1-R^2}W_2(r)\} \end{bmatrix} \quad (2.20)$$

where  $W_1(r)$  and  $W_2(r)$  are independent Brownian motions. The initial condition manifests itself via the result (see, for example, Müller and Elliott, 2003)

$$T^{-1/2}(u_y, \lfloor rT \rfloor) - u_{y,1} \Rightarrow \sigma_v K_c(r) \quad (2.21)$$

where

$$K_c(r) = \begin{cases} W_1(r) & c = 0 \\ \alpha(e^{rc} - 1)/\sqrt{-2c} + W_{1c}(r) & c < 0 \end{cases} \quad (2.22)$$

and  $W_{1c}(r)$  is the Ornstein-Uhlenbeck process

$$W_{1c}(r) = c \int_0^r e^{c(r-s)} W_1(s) ds + W_1(r). \quad (2.23)$$

The following theorem now provides the limit distributions of the three covariate augmented unit root statistics<sup>1</sup>.

**Theorem 1** *For the DGP given by (2.1)-(2.2), under Assumptions 1 and 2, with  $\rho = 1 + c/T$ ,  $c \leq 0$ ,*

*(i) For model  $i$  ( $i=A, B, C$ ),*

$$t_{\hat{\phi}} \Rightarrow \frac{c}{\sqrt{1-R^2}} \sqrt{\int_0^1 L_c^i(r)^2 dr} + \sqrt{1-R^2} \frac{\int_0^1 L_c^i(r) dW_1(r)}{\sqrt{\int_0^1 L_c^i(r)^2 dr}} - R \frac{\int_0^1 L_c^i(r)^2 dW_2(r)}{\sqrt{\int_0^1 L_c^i(r)^2 dr}} \quad (2.24)$$

where

---

<sup>1</sup>Proofs are provided in Appendix A.

$$L_c^A(r) = K_c(r) - \int_0^1 K_c(s)ds$$

$$L_c^B(r) = L_c^C(r) = K_c(r) - \{4 \int_0^1 K_c(s)ds - 6 \int_0^1 sK_c(s)ds\} - \{12 \int_0^1 sK_c(s)ds - 6 \int_0^1 K_c(s)ds\}r.$$

(ii) For model  $i$  ( $i=A, B, C$ ),

$$t_{\bar{\phi}} \Rightarrow \frac{c}{\sqrt{1-R^2}} \sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr} + \sqrt{1-R^2} \frac{\int_0^1 M_{c,\bar{c}}^i(r) dW_1(r)}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} - R \frac{\int_0^1 M_{c,\bar{c}}^i(r)^2 dW_2(r)}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} \quad (2.25)$$

$$+ \frac{c\alpha}{\sqrt{-2c(1-R^2)}} \frac{\int_0^1 M_{c,\bar{c}}^i(r) dr}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} + \frac{1}{\sqrt{1-R^2}} \frac{N_{c,\bar{c}}^i \{c \int_0^1 r M_{c,\bar{c}}^i(r) dr - \int_0^1 M_{c,\bar{c}}^i(r) dr\}}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} \quad (2.26)$$

$$+ \frac{R}{\sqrt{1-R^2}} \frac{P^i \int_0^1 M_{c,\bar{c}}^i(r) dr + Q^i \int_0^1 M_{c,\bar{c}}^i(r) dr}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} \quad (2.27)$$

where

$$M_{c,\bar{c}}^A(r) = K_c(r)$$

$$M_{c,\bar{c}}^B(r) = M_{c,\bar{c}}^C(r) = K_c(r) - \{\bar{c}^* K_c(1) + 3(1 - \bar{c}^*) \int_0^1 sK_c(s)ds\}r$$

$$N_{c,\bar{c}}^A = 0$$

$$N_{c,\bar{c}}^B = N_{c,\bar{c}}^C = \bar{c}^* K_c(1) + 3(1 - \bar{c}^*) \int_0^1 r K_c(r) dr$$

$$P^A = P_B = RW_1(1) + \sqrt{1-R^2}W_2(1)$$

$$P^C = 4\{RW_1(1) + \sqrt{1-R^2}W_2(1)\} - 6\{R \int_0^1 r dW_1(r) + \sqrt{1-R^2} \int_0^1 r dW_2(r)\}$$

$$Q^A = Q^B = 0$$

$$Q^C = 12\{R \int_0^1 r dW_1(r) + \sqrt{1-R^2} \int_0^1 r dW_2(r)\} + 6\{RW_1(1) + \sqrt{1-R^2}W_2(1)\}$$

$$\text{with } \bar{c}^* = (1 - \bar{c} + \bar{c}^2/3)^{-1}(1 - \bar{c}).$$

(iii) For model  $i$  ( $i=A, B, C$ ),

$$\hat{\Lambda} \Rightarrow G_{c,\bar{c}}^i + H_{c,\bar{c}}^i + \frac{R^2}{1-R^2}(\bar{c}^2 - 2c\bar{c}) \int_0^1 S_c^i(r)^2 dr + \frac{R^2}{\sqrt{1-R^2}} 2\bar{c} \int_0^1 S_c^i(r) dW_2(r) \quad (2.28)$$

where

$$G_{c,\bar{c}}^A = \bar{c}^2 \int_0^1 K_c(r)^2 dr - \bar{c}K_c(1)$$

$$G_{c,\bar{c}}^B = G_{c,\bar{c}}^C = \bar{c}^2 \int_0^1 K_c(r)^2 dr + (1 - \bar{c})K_c(1)^2 - k_{\bar{c}}^{-1} \{(1 - \bar{c})K_c(1) + \bar{c}^2 \int_0^1 r K_c(r) dr\}^2$$

$$H_{c,\bar{c}}^A = H_{c,\bar{c}}^C = 0$$

$$\begin{aligned} H_{c,\bar{c}}^B &= k_{\bar{c}}^{-1} \left\{ (1 - \bar{c})K_c(1) + \bar{c}^2 \int_0^1 r K_c(r) dr \right\}^2 - \left\{ k_{\bar{c}} + \frac{\bar{c}^2 R^2}{12(1-R^2)} \right\}^{-1} \times \\ & \left[ (1 - \bar{c})K_c(1) + \bar{c}^2 \int_0^1 r K_c(r) dr + \frac{R^2}{1-R^2} \left\{ \frac{\bar{c}(1-\bar{c})}{2} \int_0^1 K_c(r) dr - \bar{c}(c - \bar{c}) \int_0^1 r K_c(r) dr \right\} \right. \\ & \left. + \frac{R}{\sqrt{1-R^2}} \left\{ \bar{c} \int_0^1 r dW_2(r) - \frac{\bar{c}}{2} \int_0^1 dW_2(r) \right\} \right]^2 \\ S_c^A(r) &= S_c^B(r) = K_c(r) - \int_0^1 K_c(s) ds \\ S_c^C(r) &= K_c(r) - (4 - 6r) \int_0^1 K_c(s) ds - (12r - 6) \int_0^1 s K_c(s) ds \\ & \text{with } k_{\bar{c}} = 1 - \bar{c} + \bar{c}^2/3. \end{aligned}$$

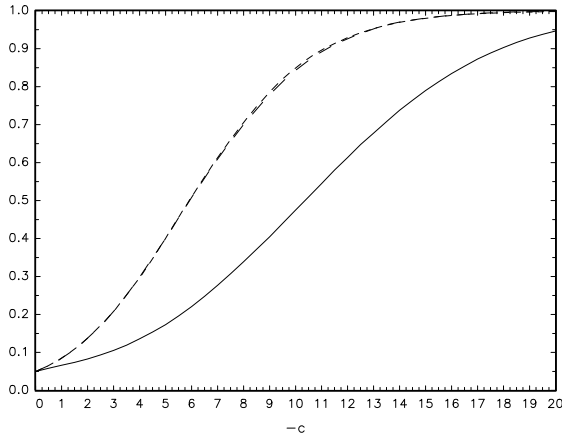
We now consider numerical results for the asymptotic properties of the tests presented so far in this chapter. In Table 2.1 we report asymptotic null (left-tail) critical values for  $R^2 = \{0, 0.1, \dots, 0.9\}$  for all tests at the nominal 0.10-, 0.05- and 0.01-levels, which were obtained by direct simulation of the limit representations of Theorem 1 with  $c = 0$  (note from (2.22) that the limits are not dependent on when  $c = 0$ ). For all asymptotic results in this chapter, we conducted Monte Carlo simulations using Gauss 9.0 with 50,000 replications, approximating the Brownian motion processes  $W_1(r)$  and  $W_2(r)$  using independent  $NIID(0, 1)$  random variates for each, and approximating the corresponding integrals by normalized sums of 2000 steps.

Table 2.1: Asymptotic $\xi$ -level critical values of covariate augmented unit root tests									
	$t_{\hat{\phi}}$			$t_{\tilde{\phi}}$			$\hat{\Lambda}$		
$R^2$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$
<i>Model A</i>									
0.0	-2.57	-2.86	-3.40	-1.61	-1.94	-2.60	4.60	3.30	1.92
0.1	-2.52	-2.82	-3.39	-1.57	-1.91	-2.57	4.80	3.36	1.67
0.2	-2.46	-2.77	-3.37	-1.52	-1.88	-2.53	5.08	3.44	1.42
0.3	-2.40	-2.72	-3.33	-1.47	-1.82	-2.51	5.45	3.60	4.22
0.4	-2.33	-2.65	-3.28	-1.41	-1.77	-2.46	5.95	3.85	1.06
0.5	-2.25	-2.58	-3.21	-1.34	-1.71	-2.41	6.64	4.28	0.98
0.6	-2.16	-2.50	-3.15	-1.27	-1.64	-2.35	7.72	4.99	1.08
0.7	-2.05	-2.40	-3.06	-1.18	-1.57	-2.29	9.61	6.25	1.52
0.8	-1.92	-2.27	-2.95	-1.07	-1.48	-2.21	13.37	8.99	2.90
0.9	-1.74	-2.10	-2.78	-0.95	-1.39	-2.16	24.82	17.64	7.97
<i>Model B</i>									
0.0	-3.13	-3.42	-3.98	-2.56	-2.85	-3.43	6.90	5.66	3.92
0.1	-3.05	-3.35	-3.90	-2.52	-2.81	-3.37	7.22	5.70	3.55
0.2	-2.98	-3.28	-3.83	-2.46	-2.77	-3.32	7.71	5.90	3.30
0.3	-2.89	-3.20	-3.76	-2.41	-2.71	-3.28	8.43	6.23	3.14
0.4	-2.79	-3.10	-3.69	-2.34	-2.65	-3.23	9.46	6.88	3.15
0.5	-2.68	-3.00	-3.59	-2.27	-2.58	-3.16	11.01	7.96	3.48
0.6	-2.54	-2.88	-3.49	-2.19	-2.50	-3.10	13.47	9.75	4.30
0.7	-2.39	-2.73	-3.36	-2.10	-2.43	-3.03	17.68	12.99	6.20
0.8	-2.20	-2.55	-3.19	-2.02	-2.34	-2.95	26.34	19.96	10.79
0.9	-1.94	-2.30	-2.97	-1.97	-2.31	-2.91	52.20	41.10	25.74

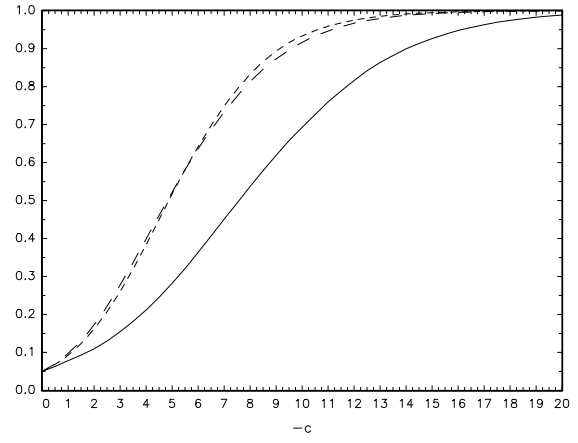
Table 2.1 (Continued): Asymptotic  $\xi$ –level critical values of covariate augmented unit root tests

	$t_{\hat{\phi}}$			$t_{\tilde{\phi}}$			$\hat{\Lambda}$		
$R^2$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$
<i>Model C</i>									
0.0	−3.13	−3.42	−3.98	−2.56	−2.85	−3.43	6.90	5.66	3.92
0.1	−3.05	−3.35	−3.90	−2.50	−2.79	−3.35	7.24	5.71	3.64
0.2	−2.98	−3.28	−3.83	−2.43	−2.74	−3.30	7.71	5.90	3.38
0.3	−2.89	−3.20	−3.76	−2.35	−2.67	−3.24	8.36	6.27	3.29
0.4	−2.79	−3.10	−3.69	−2.27	−2.59	−3.18	9.30	6.86	3.34
0.5	−2.68	−3.00	−3.59	−2.17	−2.49	−3.10	10.70	7.82	3.65
0.6	−2.54	−2.88	−3.49	−2.06	−2.39	−3.01	12.90	9.47	4.39
0.7	−2.39	−2.73	−3.36	−1.93	−2.27	−2.91	16.75	12.52	6.15
0.8	−2.20	−2.55	−3.19	−1.78	−2.14	−2.81	24.79	18.97	10.61
0.9	−1.94	−2.30	−2.97	−1.60	−2.00	−2.70	49.76	39.30	25.13

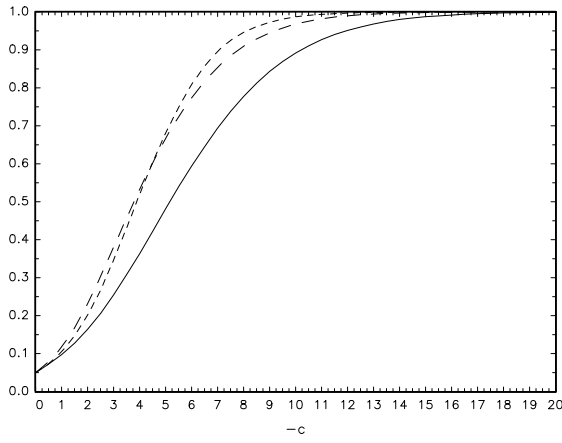
To gain some insight into the relative power performance of the three tests  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$ , we first abstract from the effect of the initial condition by making the usual assumption that it is asymptotically negligible. The limit distributions of the statistics are then as given in Theorem 1 on setting  $\alpha = 0$ . Figures 2.1-2.3 show the local asymptotic powers of the tests conducted at the nominal 0.05-level as functions  $-c = \{0, 0.5, \dots, 20.0\}$  (with  $c = 0$  corresponding to asymptotic size, i.e. 0.05) for Models A, B and C, respectively, for  $R^2 = \{0.2, 0.4, 0.6, 0.8\}$ .



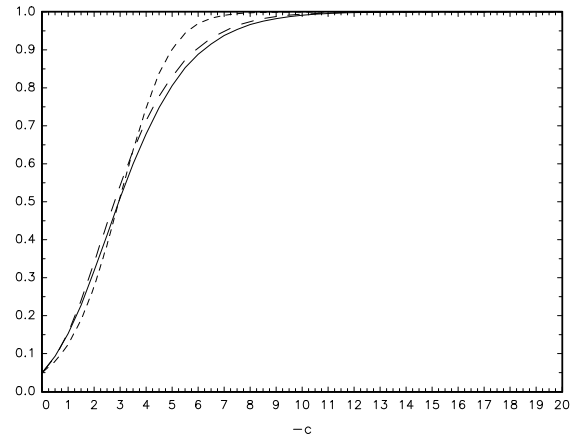
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



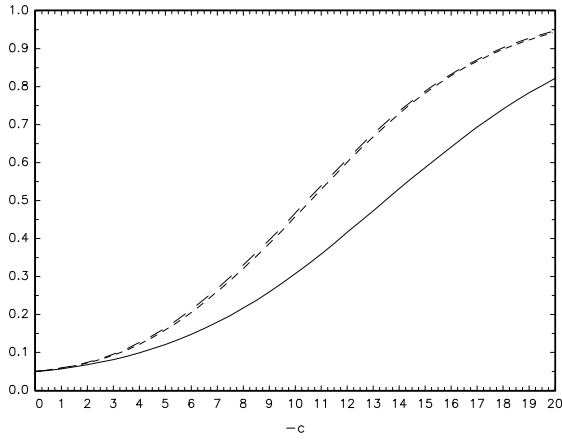
(c)  $R^2 = 0.6$



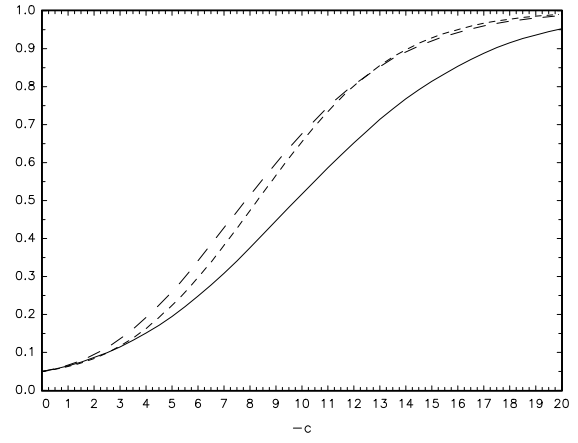
(d)  $R^2 = 0.8$

Figure 2-1: Local asymptotic power of nominal 0.05-level tests: Model A,  $\alpha = 0$ ;

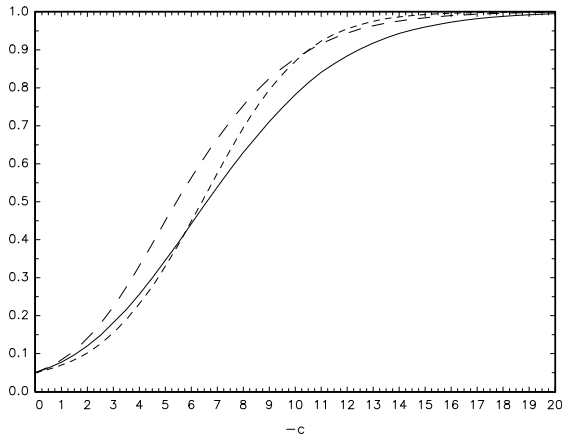
$$t_{\hat{\phi}}: \text{---} \quad t_{\tilde{\phi}}: \text{-- --} \quad \hat{\Lambda}: \text{. . .}$$



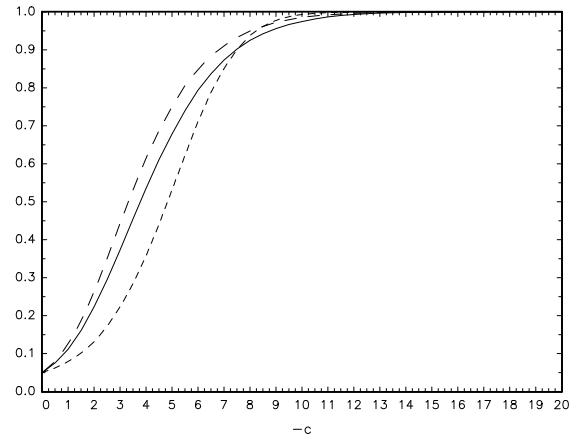
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



(c)  $R^2 = 0.6$

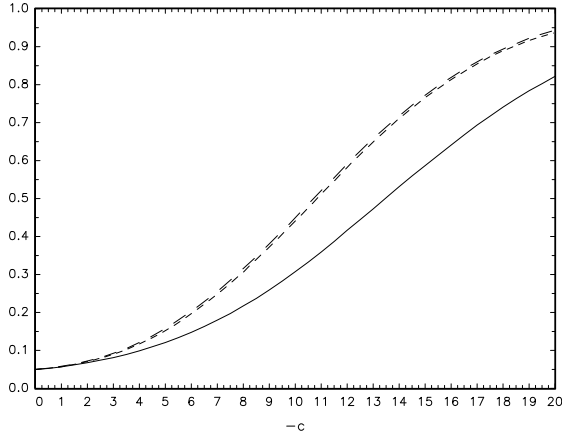


(d)  $R^2 = 0.8$

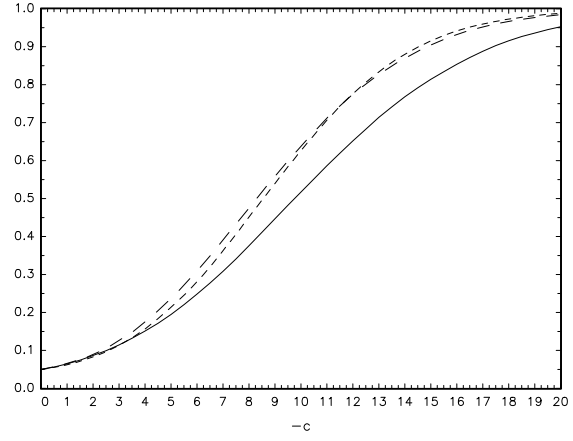
Figure 2-2: Local asymptotic power of nominal 0.05-level tests: Model B,  $\alpha = 0$ ;

$$t_{\hat{\phi}}: \text{---} \quad t_{\tilde{\phi}}: \text{-- --} \quad \hat{\Lambda}: \text{. . .}$$

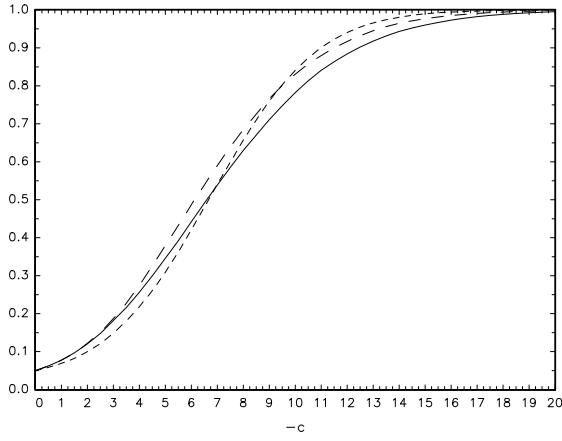




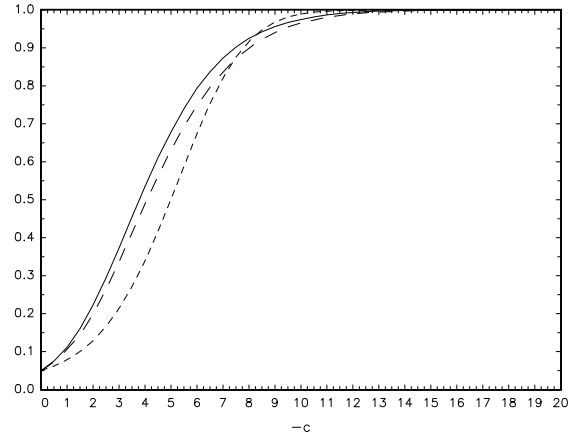
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



(c)  $R^2 = 0.6$



(d)  $R^2 = 0.8$

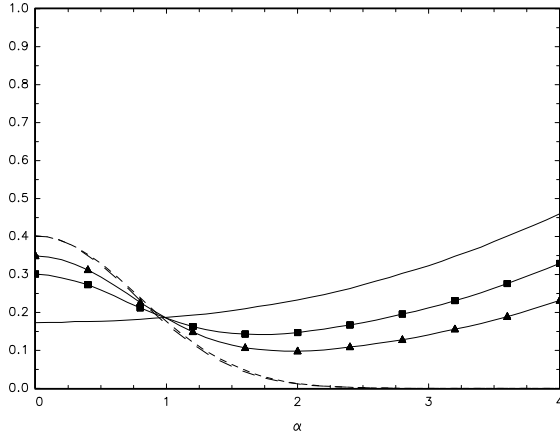
Figure 2-3: Local asymptotic power of nominal 0.05-level tests: Model C,  $\alpha = 0$ ;

$$t_{\hat{\phi}}: \text{—} \quad t_{\tilde{\phi}}: \text{--} \quad \hat{\Lambda}: \text{---}$$

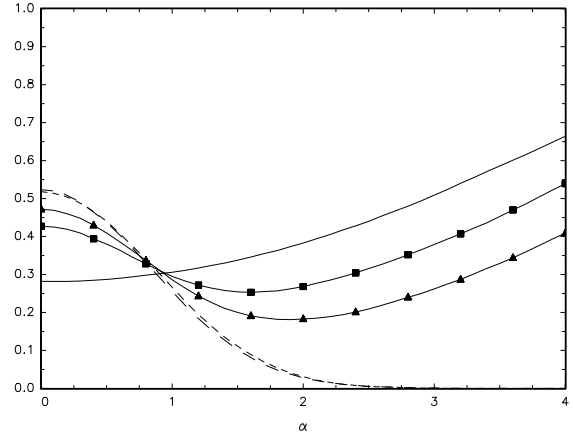
Consider first the results for Model A in Figure 2.1. We observe that for smaller values of  $R^2$ , the familiar result of the GLS-based tests delivering a substantial power advantage relative to the OLS-based test  $t_{\hat{\phi}}$  is borne out. These power advantages, however, diminish as  $R^2$  increases, so that by  $R^2 = 0.8$ , there is considerably less difference between the power profiles of the three tests. For the two GLS-based tests, there is very little to choose between them for small to moderate  $R^2$ , while for larger  $R^2$ , the  $\hat{\Lambda}$  test has slightly lower power than  $t_{\tilde{\phi}}$  for small  $-c$  but modestly higher power for some larger  $-c$ . In Figure 2.2 (Model B), as expected we see a reduction in power of all tests relative to Model A, brought about by the allowance of a trend in  $y_t$ . However, it is still the case that for small  $R^2$ , the GLS-based tests outperform  $t_{\hat{\phi}}$ , and have similar levels of

power to each other. Interestingly, as  $R^2$  increases,  $t_{\tilde{\phi}}$  becomes generally more powerful than  $\hat{\Lambda}$ , and by  $R^2 = 0.8$ , the relative power levels of  $\hat{\Lambda}$  have reduced to values generally below those of the OLS-based test  $t_{\hat{\phi}}$ . In contrast,  $t_{\tilde{\phi}}$  retains a power advantage over  $t_{\hat{\phi}}$  for all  $R^2$  considered. Finally, for Model C (Figure 2.3), similar comments apply as were made for Model B in Figure 2.2, with the additional detrending of  $x_t$  causing relatively little change to the powers of the tests. One noteworthy distinction between the results for Figures 2.2 and 2.3 is that for  $R^2 = 0.8$ , for Model C we now find that the power of the  $t_{\tilde{\phi}}$  test also drops slightly below that of  $t_{\hat{\phi}}$ , so that here the OLS-based  $t_{\hat{\phi}}$  test generally outperforms both GLS-based tests. Overall, however, from the results of Figures 2.1-2.3 we conclude that, if we abstract from potentially non-negligible initial conditions, on balance the test  $t_{\tilde{\phi}}$  offers arguably the most appealing power profile of the tests considered.

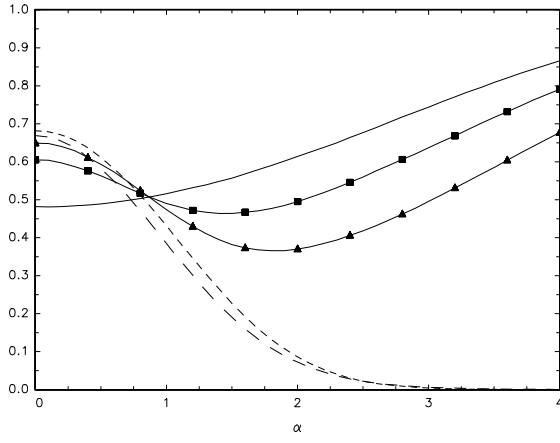
We now examine the effects of an asymptotically non-negligible initial condition on local asymptotic power. Figures 2.4-2.6 report local asymptotic powers of the five tests conducted at the nominal 0.05-level, as functions of  $\alpha = \{0, 0.1, \dots, 4.00\}$  ( $\alpha = 0$  corresponding to an asymptotically negligible initial condition); note that replacing  $\alpha$  with  $-\alpha$  would give the same results. Figure 2.4 presents results for Model A, where a representative local alternative setting of  $c = -5$  is used. Figures 2.5 and 2.6 give results for Models B and C, respectively, using  $c = -10$ .



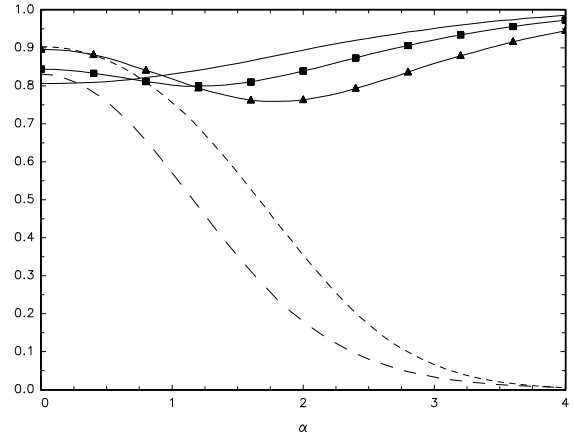
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



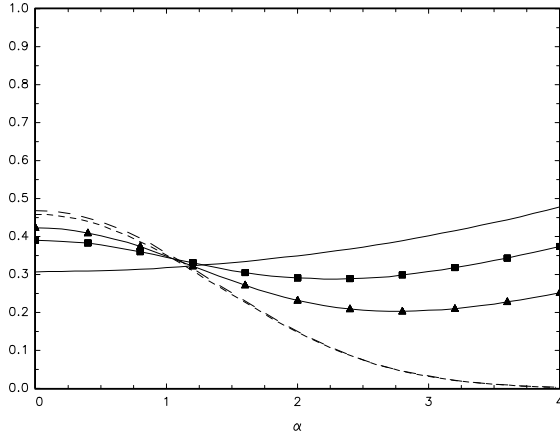
(c)  $R^2 = 0.6$



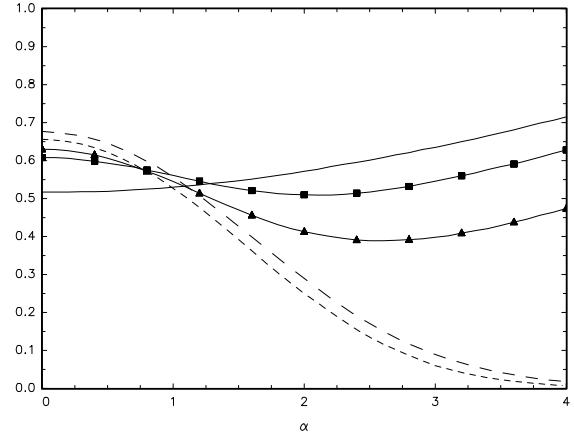
(d)  $R^2 = 0.8$

Figure 2-4: Local asymptotic power of nominal 0.05-level tests: Model A,  $c = -5$ ;

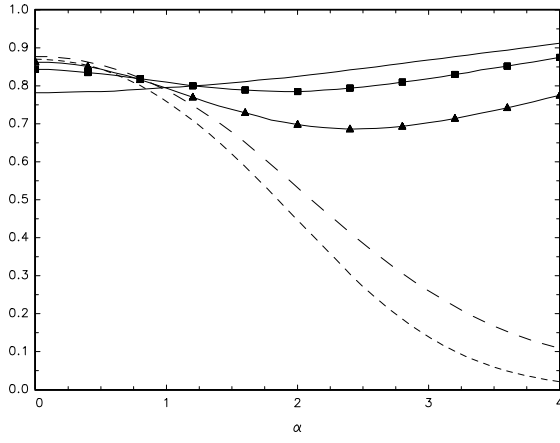
$$t_{\hat{\phi}}: \text{---} \quad t_{\tilde{\phi}}: \text{---} \quad \hat{\Lambda}: \text{---} \quad \text{---}, \quad UR(t_{\hat{\phi}}, t_{\tilde{\phi}}): \text{---}\blacksquare\text{---}, \quad UR(t_{\hat{\phi}}, \hat{\Lambda}): \text{---}\blacktriangle\text{---}$$



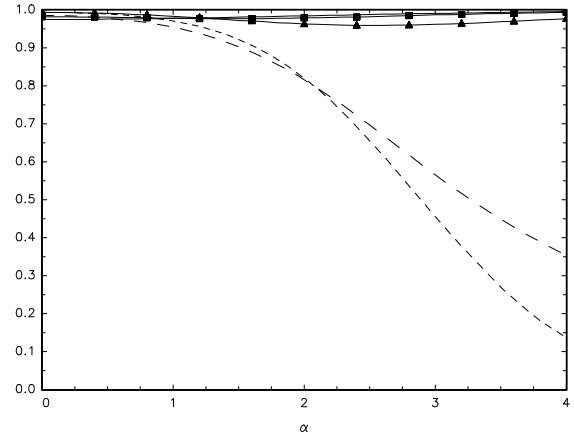
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



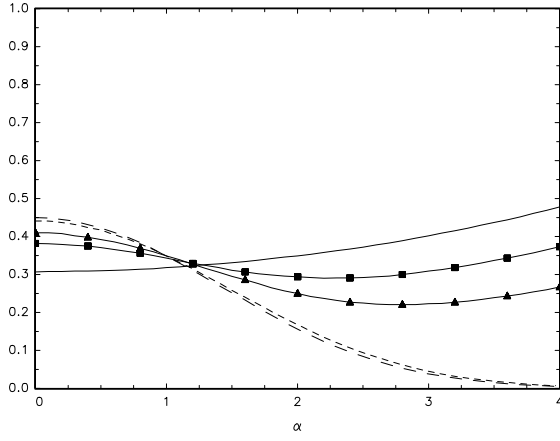
(c)  $R^2 = 0.6$



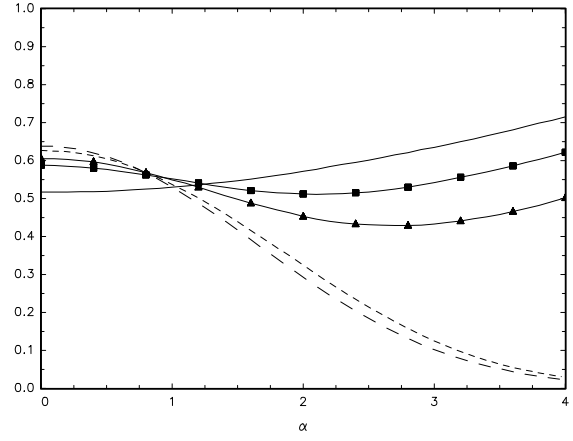
(d)  $R^2 = 0.8$

Figure 2-5: Local asymptotic power of nominal 0.05-level tests: Model B,  $c = -10$ ;

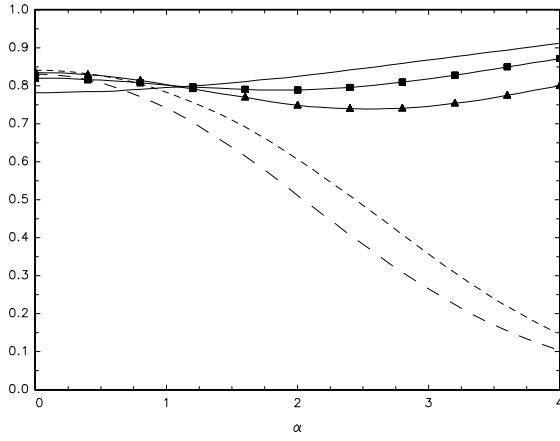
$$t_{\hat{\phi}}: \text{---} \quad t_{\tilde{\phi}}: \text{-- --} \quad \hat{\Lambda}: \text{- - -}, \quad UR(t_{\hat{\phi}}, t_{\tilde{\phi}}): \text{---}\blacksquare\text{---}, \quad UR(t_{\hat{\phi}}, \hat{\Lambda}): \text{---}\blacktriangle\text{---}$$



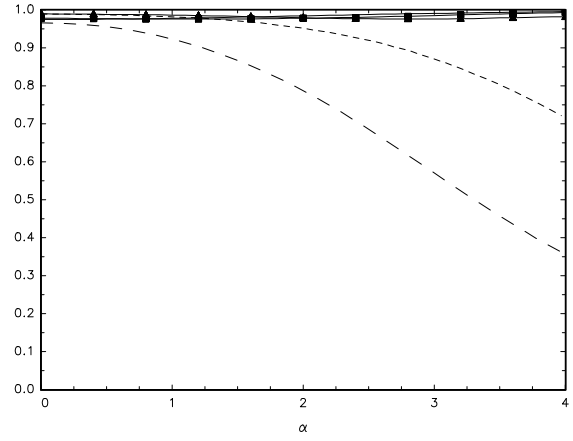
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



(c)  $R^2 = 0.6$



(d)  $R^2 = 0.8$

Figure 2-6: Local asymptotic power of nominal 0.05-level tests: Model A,  $c = -10$ ;

$$t_{\hat{\phi}}: \text{—} \quad t_{\tilde{\phi}}: \text{-- --} \quad \hat{\Lambda}: \text{---} \quad , \quad UR(t_{\hat{\phi}}, t_{\tilde{\phi}}): \text{—}\blacksquare\text{—} \quad , \quad UR(t_{\hat{\phi}}, \hat{\Lambda}): \text{—}\blacktriangle\text{—}$$

Considering first Model A in Figure 2.4, the stand out feature of these power curves is that while the power of the OLS-based  $t_{\hat{\phi}}$  test is increasing in the magnitude of the initial condition  $\alpha$ , the powers of both GLS-based tests decrease to zero as  $\alpha$  increases. Hence, while the GLS tests are more powerful for  $\alpha = 0$  (cf. Figure 2.1) and for small values of  $\alpha$ , they are much less powerful than  $t_{\hat{\phi}}$  for larger initial conditions. This pattern of results closely mirrors what is found when analysing the effects of initial conditions on standard, non-covariate augmented OLS and GLS demeaned/detrended unit root tests, and highlights the fact that GLS-based unit root tests do not deliver reliable unit root test inference in the presence of large initial conditions. Between the two GLS-based tests, there is little difference between the power profiles for  $R^2 = 0.2$  and

$R^2 = 0.4$ , while as  $R^2$  increases to 0.6 and then 0.8, it is clear that  $t_{\hat{\phi}}$  emerges as the more powerful procedure. For Model B (Figure 2.5), we observe the same broad patterns of results vis-à-vis the power of  $t_{\hat{\phi}}$  compared to the GLS-based tests. Once again,  $t_{\hat{\phi}}$  has power that increases in  $\alpha$ , while the GLS-based tests have higher power when  $\alpha = 0$ , but then a decreasing power profile as  $\alpha$  rises. Between the GLS-based tests, in contrast to Model A, here we see that  $t_{\hat{\phi}}$  offers generally the best levels of power across  $\alpha$ , particularly for larger  $\alpha$  and  $R^2$  values. Finally, the results for Model C in Figure 2.6 again highlight the general result that the GLS-based tests are typically more powerful than  $t_{\hat{\phi}}$  for zero and small  $\alpha$ , while this ranking reverses for larger  $\alpha$ . Of the two GLS-based tests, once again the powers are very similar for  $R^2 = 0.2$  and  $R^2 = 0.4$ , while for larger  $R^2$ ,  $\hat{\Lambda}$  is the better performing test.

## 2.4 A union of rejections strategy

The results of the previous section demonstrate that when the initial condition is small, we would want to apply one of the two GLS-based tests ( $t_{\hat{\phi}}$  or  $\hat{\Lambda}$ ); on the other hand, when the initial condition is larger, applying such a test would result in a (potentially substantial) loss of power relative to applying the OLS-based test  $t_{\hat{\phi}}$ . In practice, given uncertainty regarding the magnitude of the initial condition, we wish to have available a procedure that capitalises on both the relatively high power of the GLS approach when is small, and the relatively high power of  $t_{\hat{\phi}}$  otherwise. A similar issue arises in the case of non-covariate augmented unit root testing, and the approach proposed by Harvey et al. (2009) is to take a union of rejections of the OLS- and GLS-based tests, whereby the null hypothesis is rejected if either of the individual tests rejects. In the present context, this implies taking a union of rejections between  $t_{\hat{\phi}}$  and either one of the GLS-based tests.

We now set out the union of rejections approach based on  $t_{\hat{\phi}}$  (here denoted by  $t_{OLS}$ ) and a GLS-based test ( $t_{\hat{\phi}}$  or  $\hat{\Lambda}$ ) denoted by  $t_{GLS}$ . Denoting the asymptotic  $\xi$ -level critical values of these tests by  $cv_{\xi}^{OLS}$  and  $cv_{\xi}^{GLS}$ , respectively, we can define the simple union of rejections strategy by the decision rule

$$\text{Reject } H_0 \text{ if } \{t_{OLS} < cv_{\xi}^{OLS} \text{ or } t_{GLS} < cv_{\xi}^{GLS}\}. \quad (2.29)$$

An alternative way of representing this decision rule is to express it in terms a single test statistic,

$t_{UR}$ , as follows:

$$\text{Reject } H_0 \text{ if } \left\{ t_{UR} = \min \left( t_{OLS}, \frac{cv_{\xi}^{OLS}}{cv_{\xi}^{GLS}} t_{GLS} \right) < cv_{\xi}^{OLS} \right\}. \quad (2.30)$$

If we use  $\mathcal{L}^{OLS}$  and  $\mathcal{L}^{GLS}$  to denote the generic joint limit distributions of  $t_{OLS}$  and  $t_{GLS}$ , respectively (i.e. the right-hand-side expressions given in Theorem 1), an application of the continuous mapping theorem establishes that

$$t_{UR} \Rightarrow \min \left( \mathcal{L}^{OLS}, \frac{cv_{\xi}^{OLS}}{cv_{\xi}^{GLS}} \mathcal{L}^{GLS} \right). \quad (2.31)$$

The Bonferroni bound for the asymptotic size of this procedure under the null is  $2\xi$  (since it simply involves rejecting the null when either of the individual tests reject). Harvey et al. (2009) suggest restoring the union of rejections asymptotic size to the nominal level  $\xi$  by applying a common positive scaling constant,  $\psi_{\xi} > 1$ , to the (negative) critical values  $cv_{\xi}^{OLS}$  and  $cv_{\xi}^{GLS}$  (so that  $t_{OLS}$  is compared with  $\psi_{\xi} cv_{\xi}^{OLS}$  and  $t_{GLS}$  with  $\psi_{\xi} cv_{\xi}^{GLS}$ ), such that in the limit, rejection of the null occurs with probability  $\xi$ .

While this approach extends naturally to the covariate augmented unit root testing problem when using  $t_{\hat{\phi}}$  for  $t_{GLS}$ , since here both  $cv_{\xi}^{OLS}$  and  $cv_{\xi}^{GLS}$  are negative, we cannot apply such a simple adjustment when using  $\hat{\Lambda}$ , since the latter test has positive critical values, and scaling by  $\psi_{\xi} > 1$  would induce greater asymptotic size. An adjustment that is applicable in all cases is to first apply a common additive adjustment to both  $t_{GLS}$  and  $cv_{\xi}^{GLS}$ , say  $t_{GLS} - \lambda_{\xi}$  and  $cv_{\xi}^{GLS} - \lambda_{\xi}$ , such that the GLS-based test decision rule is unchanged, but that the adjusted critical value  $cv_{\xi}^{GLS} - \lambda_{\xi}$  equals  $cv_{\xi}^{OLS}$ , i.e.  $\lambda_{\xi} = cv_{\xi}^{GLS} - cv_{\xi}^{OLS}$ . Once the critical values are lined up in this way, the critical values are both negative, and a Harvey et al. (2009)-type multiplicative scaling can be applied to control the asymptotic size. More formally we propose the following union of rejections decision rule:

$$\text{Reject } H_0 \text{ if } \{ t_{OLS} < \psi_{\xi} cv_{\xi}^{OLS} \text{ or } t_{GLS} - \lambda_{\xi} < \psi_{\xi} cv_{\xi}^{OLS} \} \quad (2.32)$$

or, equivalently,

$$\text{Reject } H_0 \text{ if } \{ t_{UR}^* = \min(t_{OLS}, t_{GLS} - \lambda_{\xi}) < \psi_{\xi} cv_{\xi}^{OLS} \}. \quad (2.33)$$

In the limit we obtain

$$t_{UR}^* \Rightarrow \min(\mathcal{L}^{OLS}, \mathcal{L}^{GLS} - \lambda_{\xi}) \quad (2.34)$$

and we compute  $\psi_\xi$  by simulation of the limit distribution of  $t_{UR}^*$ , calculating the  $\xi$ -level null critical value for this distribution, say  $cv_\xi^{UR}$ , and then evaluate  $\psi_\xi$  as  $\psi_\xi = cv_\xi^{UR}/cv_\xi^{OLS}$ . In what follows, we consider two union of rejections procedures, one based on a union of  $t_{\hat{\phi}}$  and  $t_{\tilde{\phi}}$ , the other based on a union of  $t_{\hat{\phi}}$  and  $\hat{\Lambda}$ ; hereafter we denote these unions by  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  respectively. Values for  $\psi_\xi$  for  $R^2 = \{0, 0.1, \dots, 0.9\}$  at the nominal 0.10-, 0.05- and 0.01-levels are given in Table 2.2 for each of these union of rejections strategies.



Table 2.2: Asymptotic  $\psi_\xi$  values for  $\xi$ -level union of rejections procedures

	$UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$			$UR(t_{\hat{\phi}}, \hat{\Lambda})$		
$R^2$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$
<i>Model A</i>						
0.0	1.099	1.081	1.062	1.200	1.132	1.071
0.1	1.105	1.083	1.057	1.212	1.143	1.080
0.2	1.108	1.086	1.056	1.226	1.157	1.090
0.3	1.111	1.091	1.057	1.240	1.168	1.099
0.4	1.113	1.093	1.064	1.261	1.184	1.106
0.5	1.119	1.097	1.068	1.276	1.205	1.117
0.6	1.126	1.102	1.070	1.297	1.223	1.135
0.7	1.135	1.106	1.073	1.336	1.236	1.149
0.8	1.152	1.115	1.076	1.389	1.271	1.172
0.9	1.179	1.136	1.082	1.502	1.352	1.207
<i>Model B</i>						
0.0	1.063	1.053	1.039	1.134	1.107	1.058
0.1	1.063	1.055	1.042	1.148	1.115	1.064
0.2	1.065	1.055	1.041	1.162	1.128	1.082
0.3	1.065	1.055	1.039	1.179	1.130	1.091
0.4	1.066	1.057	1.040	1.197	1.149	1.098
0.5	1.066	1.058	1.047	1.220	1.160	1.107
0.6	1.068	1.061	1.045	1.248	1.178	1.111
0.7	1.070	1.060	1.045	1.285	1.204	1.133
0.8	1.075	1.060	1.044	1.364	1.261	1.152
0.9	1.088	1.067	1.047	1.499	1.362	1.205

Table 2.2 (Continued): Asymptotic $\psi_\xi$ values for $\xi$ -level union of rejections procedures						
	$UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$			$UR(t_{\hat{\phi}}, \hat{\Lambda})$		
$R^2$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$
<i>Model C</i>						
0.0	1.063	1.053	1.039	1.134	1.107	1.058
0.1	1.065	1.055	1.042	1.147	1.113	1.067
0.2	1.067	1.056	1.040	1.157	1.119	1.074
0.3	1.069	1.056	1.042	1.169	1.130	1.085
0.4	1.072	1.061	1.041	1.183	1.135	1.095
0.5	1.075	1.063	1.048	1.197	1.144	1.102
0.6	1.080	1.065	1.047	1.222	1.155	1.107
0.7	1.085	1.071	1.048	1.251	1.186	1.121
0.8	1.095	1.075	1.049	1.318	1.235	1.147
0.9	1.126	1.093	1.061	1.491	1.353	1.202

The union of rejection strategies  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  are by construction asymptotically correctly sized. We now consider the asymptotic local power properties of  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  in relation to the powers of the individual tests, the results for which are also displayed in Figures 2.4-2.6. Consider first Model A in Figure 2.4, and to aid comparison of the union of rejections procedures, consider an informal (and infeasible) power “envelope” formed from the limit power of  $\hat{\Lambda}$  for values of  $\alpha$  up to the point where  $\hat{\Lambda}$  and  $t_{\hat{\phi}}$  have the same power, and from the limit power of  $t_{\hat{\phi}}$  for  $\alpha$  beyond this point. With reference to this envelope, both  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  do a decent job of tracking its broad shape, offering decent power levels across the range of  $\alpha$  values considered. Both  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  capture much of the power advantage that  $\hat{\Lambda}$  holds over  $t_{\hat{\phi}}$  for small  $\alpha$ , while also achieving the substantial power gain that  $t_{\hat{\phi}}$  holds over either of the GLS-based tests for larger  $\alpha$ , with power profiles that are increasing in  $\alpha$  as opposed to approaching zero. Of the two union of rejections procedures, a trade-off clearly exists between the higher power  $\alpha$  for small that  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  achieves, and the higher power for larger  $\alpha$  that  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  displays.

However, in the small- $\alpha$  region where  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  outperforms  $UR(t_{\hat{\phi}}, t_{\hat{\phi}})$ , the power differences are relatively modest, while in the larger- $\alpha$  range where  $UR(t_{\hat{\phi}}, t_{\hat{\phi}})$  outperforms  $UR(t_{\hat{\phi}}, \hat{\Lambda})$ , the gains can be quite substantial. For this reason, we consider that  $UR(t_{\hat{\phi}}, t_{\hat{\phi}})$  arguably offers the more preferable power profile of the two procedures overall. For Models B and C in Figures 2.5 and 2.6, broadly similar comments can be applied, with both  $UR(t_{\hat{\phi}}, t_{\hat{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  tracking the shape of the informal envelope comprised of the best performing tests for each region of  $\alpha$ . What is noticeable is that, compared to Model A, the power differences between  $UR(t_{\hat{\phi}}, t_{\hat{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  are less marked for small  $\alpha$ , yet still quite substantial for large  $\alpha$ , adding weight to the argument for our preference for  $UR(t_{\hat{\phi}}, t_{\hat{\phi}})$  over  $UR(t_{\hat{\phi}}, \hat{\Lambda})$ .

## 2.5 Finite sample performance

In this section we consider the finite sample behaviour of the individual tests of section 2.2 and the proposed union of rejections procedures  $UR(t_{\hat{\phi}}, t_{\hat{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  under Assumptions 1 and 2. In order to implement the tests in such a setting, we first require a consistent estimator of  $R^2$ , given that all the tests have critical values that depend on this unknown quantity (the union of rejections procedures also require  $R^2$ -dependent scaling values). Under our assumptions, the estimator

$$\hat{R}^2 = \frac{\hat{\sigma}_{ev}^2}{\hat{\sigma}_v^2 \hat{\sigma}_e^2} \quad (2.35)$$

where

$$\hat{\sigma}_v^2 = T^{-1} \sum_{t=2}^T \hat{v}_t^2, \quad \hat{\sigma}_e^2 = T^{-1} \sum_{t=1}^T \hat{e}_t^2, \quad \hat{\sigma}_{ev} = T^{-1} \sum_{t=2}^T \hat{e}_t \hat{v}_t \quad (2.36)$$

with  $\hat{e}_t = \hat{u}_{x,t}$  and  $\hat{v}_t$  being the residual from a regression of  $\hat{u}_{y,t}$  on  $\hat{u}_{y,t-1}$  can be shown to provide a consistent estimator of  $R^2$ . Additionally, as highlighted in section 2.2.2, the  $\hat{\Lambda}$  statistic requires a consistent estimator of  $\Omega$ . Given that  $\Omega$  is also comprised of  $\sigma_v^2$ ,  $\sigma_e^2$  and  $\sigma_{ev}$  a natural estimator is to use

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_v^2 & \hat{\sigma}_{ev} \\ \hat{\sigma}_{ev} & \hat{\sigma}_e^2 \end{bmatrix} \quad (2.37)$$

which can be shown to be consistent for  $\Omega$ . Note that both  $\hat{R}^2$  and  $\hat{\Omega}$  remain consistent when  $\alpha$  in Assumption 2 is not equal to zero; in contrast, the corresponding estimators outlined in Elliott and Jansson (2003, p.81) are only consistent under the local-to-unit root alternative when the initial

condition is asymptotically negligible (i.e.  $\alpha = 0$ ), due to their reliance on a first difference-based (cf. GLS-based) demeaning/detrending of  $y_t$ .

Our Monte Carlo simulations are based on generating (2.1)-(2.2) for  $T = 150$  using 50,000 replications, with  $\varepsilon_t = \begin{bmatrix} v_t & e_t \end{bmatrix}' \sim IIN(0, \Omega)$ ,  $\sigma_v^2 = \sigma_e^2 = 1$ ,  $\sigma_{ev} = R^2 = \{0.2, 0.4, 0.6, 0.8\}$ , and with  $\mu_y = \beta_y = \mu_x = \beta_x = 0$ . We first simulated the empirical size of the  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$ ,  $\hat{\Lambda}$  tests and the  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$ ,  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  procedures at the nominal 0.05-level, setting  $\rho = 1$  in (2.2). Asymptotic critical values were used, linearly interpolating between the values in Tables 2.1 and 2.2 on the basis of  $\hat{R}^2$ . The results for Models A, B and C are reported in Table 2.3, and we observe only modest finite sample size distortions across the different tests and values of  $R^2$ . For larger  $R^2$  in the case of Models B and C,  $\hat{\Lambda}$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  are a little under-sized, while  $t_{\tilde{\phi}}$  and  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  are a little over-sized for all cases. However, all sizes for  $t_{\tilde{\phi}}$  and  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  are below 0.07 and 0.06 respectively, hence finite sample size distortion does not appear to be a major concern for these procedures.

Table 2.3: Finite sample size of nominal 0.05-level covariate augmented unit root tests: $T = 150$					
$R^2$	$t_{\hat{\phi}}$	$t_{\tilde{\phi}}$	$\hat{\Lambda}$	$UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$	$UR(t_{\hat{\phi}}, \hat{\Lambda})$
<i>Model A</i>					
0.2	0.053	0.064	0.053	0.059	0.053
0.4	0.053	0.065	0.051	0.057	0.050
0.6	0.050	0.064	0.048	0.056	0.048
0.8	0.048	0.061	0.041	0.054	0.042
<i>Model B</i>					
0.2	0.054	0.064	0.049	0.059	0.050
0.4	0.053	0.063	0.047	0.057	0.049
0.6	0.052	0.060	0.043	0.055	0.044
0.8	0.048	0.053	0.029	0.049	0.032
<i>Model C</i>					
0.2	0.054	0.064	0.047	0.059	0.049
0.4	0.053	0.063	0.045	0.058	0.047
0.6	0.051	0.062	0.040	0.056	0.043
0.8	0.047	0.059	0.027	0.053	0.030

Of most interest are the relative finite sample powers of the procedures, and Figures 2.7, 2.8 and 2.9 present results for Models A, B and C, for settings that correspond to the local asymptotic power results in Figures 2.4, 2.5 and 2.6. Here, we set  $\rho = 1 + c/T$  with  $c = -5$  for Model A (Figure 2.7) and  $c = -10$  for Models B and C (Figures 2.8 and 2.9), and report the estimated powers of nominal 0.05-level tests across  $\alpha = \{0., 0.1, \dots, 4.0\}$ . In each case, we find that the relative finite sample powers bear a very close resemblance to the corresponding local asymptotic results, with the powers of  $t_{\hat{\phi}}$  increasing in  $\alpha$ , the powers of  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  initially higher than that for  $t_{\hat{\phi}}$  for small  $\alpha$ , but then falling towards zero as  $\alpha$  increases, and the  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  procedures capturing a proportion of the higher  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  power for small  $\alpha$ , and a proportion of the higher power of  $t_{\hat{\phi}}$

for larger  $\alpha$ . Compared to the asymptotic results,  $t_{\hat{\phi}}$  and  $UR(t_{\hat{\phi}}, t_{\hat{z}})$  appear to have higher relative power for  $T = 150$ , which arises as a result of the small over-size seen for these procedures, but otherwise the finite sample and large sample results are very similar. What is clear is that the union of rejections procedures offer robust power profiles across the full range of initial conditions, avoiding the low power that can arise from use of the GLS-based tests alone while retaining a good proportion of the additional power offered by the GLS-based tests over the OLS-based variant. Of the two,  $UR(t_{\hat{\phi}}, t_{\hat{z}})$  emerges as the test with arguably the most attractive power properties overall, and on the basis of both the asymptotic and finite sample results, it is this procedure that we recommend for practical applications.

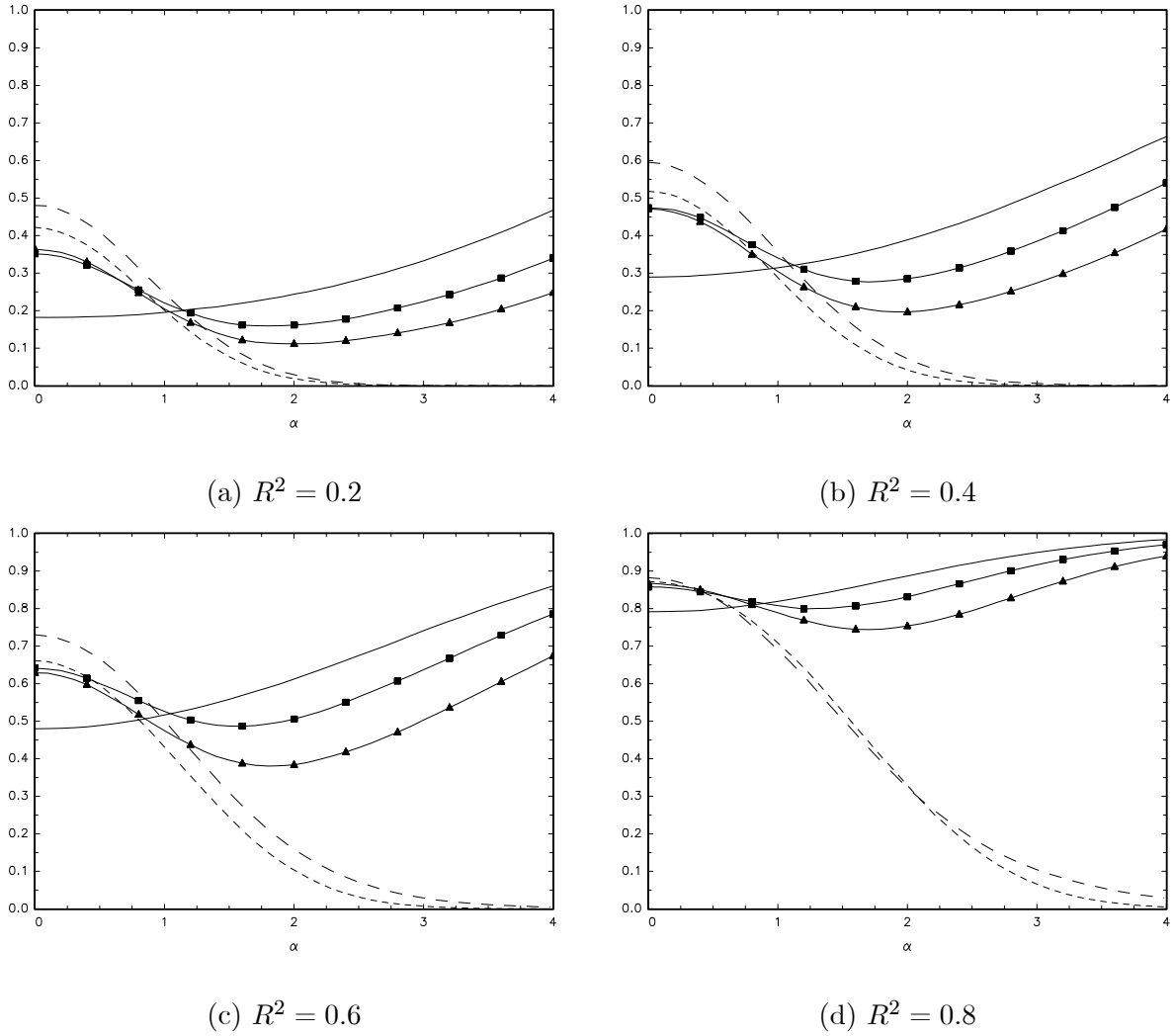
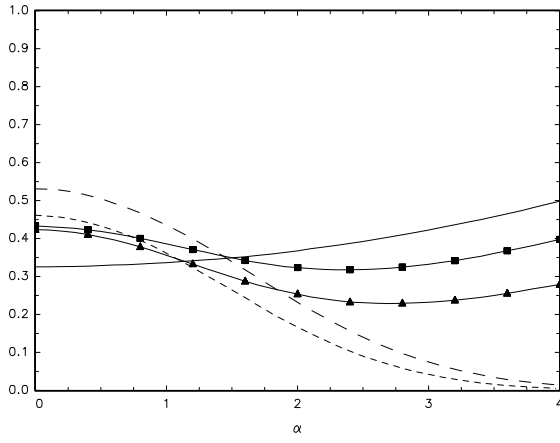
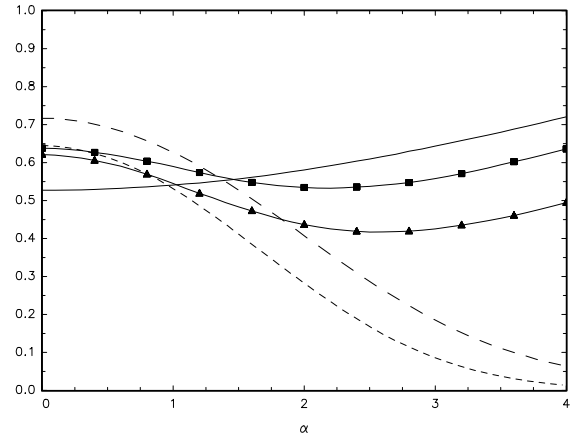


Figure 2-7: Finite sample power of nominal 0.05-level tests: Model A,  $T = 150$ ,  $c = -5$ ;

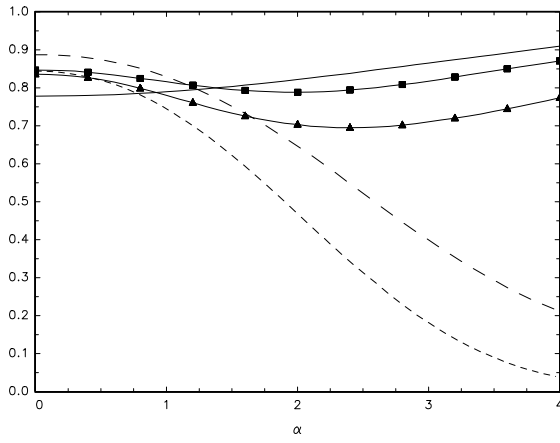
$$t_{\hat{\phi}}: \text{—} \quad t_{\hat{z}}: \text{-- --} \quad \hat{\Lambda}: \text{...}, \quad UR(t_{\hat{\phi}}, t_{\hat{z}}): \text{—}\blacksquare\text{—}, \quad UR(t_{\hat{\phi}}, \hat{\Lambda}): \text{—}\blacktriangle\text{—}$$



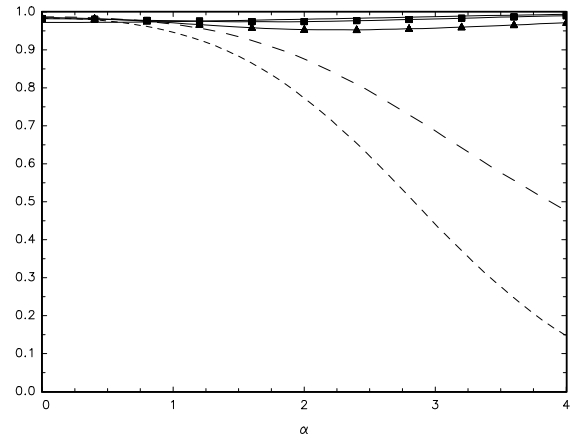
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



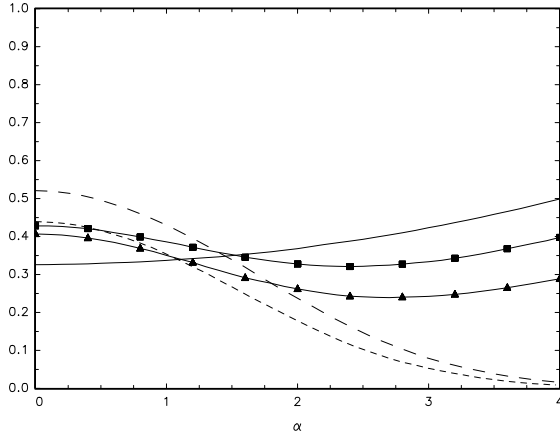
(c)  $R^2 = 0.6$



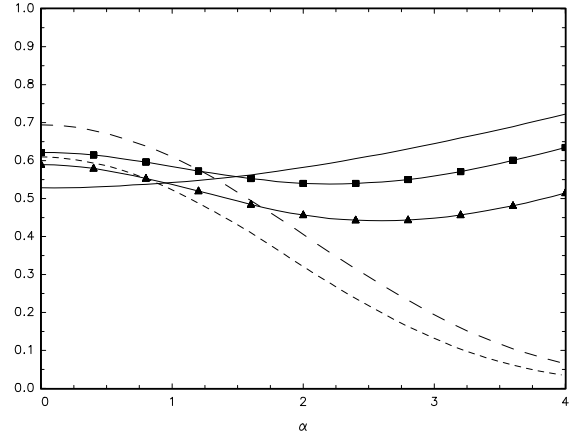
(d)  $R^2 = 0.8$

Figure 2-8: Finite sample power of nominal 0.05-level tests: Model B,  $T = 150$ ,  $c = -10$ ;

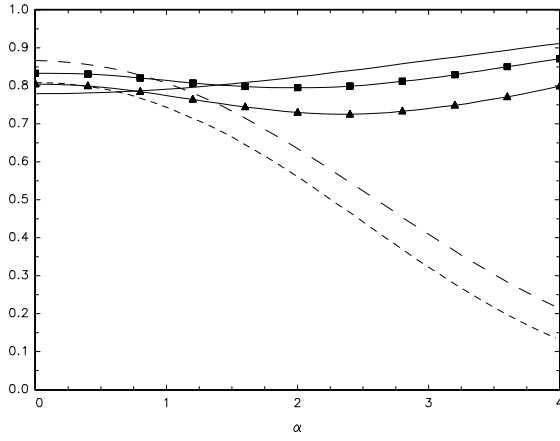
$$t_{\hat{\phi}}: \text{---} \quad t_{\tilde{\phi}}: \text{-- --} \quad \hat{\Lambda}: \text{---} \quad UR(t_{\hat{\phi}}, t_{\tilde{\phi}}): \text{---}\blacksquare\text{---} \quad , \quad UR(t_{\hat{\phi}}, \hat{\Lambda}): \text{---}\blacktriangle\text{---}$$



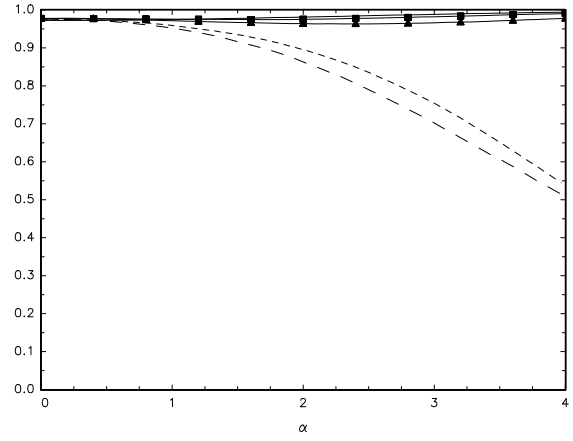
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



(c)  $R^2 = 0.6$



(d)  $R^2 = 0.8$

Figure 2-9: Finite sample power of nominal 0.05-level tests: Model C,  $T = 150$ ,  $c = -10$ ;

$$t_{\hat{\phi}}: \text{---} \quad t_{\tilde{\phi}}: \text{-- --} \quad \hat{\Lambda}: \text{---} \quad UR(t_{\hat{\phi}}, t_{\tilde{\phi}}): \text{---}\blacksquare\text{---} \quad , \quad UR(t_{\hat{\phi}}, \hat{\Lambda}): \text{---}\blacktriangle\text{---}$$

In practice, when implementing  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  we would want to allow for additional serial correlation in  $u_{y,t}$  and  $u_{x,t}$ . In order to admit more general serial correlation into our DGP, we consider the following simple autoregressive-based extension to (2.2):

$$\begin{bmatrix} a(L)(u_{y,t} - \rho u_{y,t-1}) \\ b(L)u_{x,t} \end{bmatrix} = \begin{bmatrix} v_t \\ e_t \end{bmatrix} \quad (2.38)$$

with

$$a(L) = 1 - a_1 L - \dots - a_p L^p, \quad (2.39)$$

$$b(L) = 1 - b_1 L - \dots - b_q L^q \quad (2.40)$$



where the roots of  $a(L)$  and  $b(L)$  all lie outside the unit circle, and where  $\varepsilon_t = \begin{bmatrix} v_t & e_t \end{bmatrix}'$  continues to satisfy Assumption 1. We also modify Assumption 2 so that, when  $c < 0$ ,  $u_{y,1} = \alpha\sqrt{\omega^2/(1-\rho^2)}$ , where  $\omega^2$  denotes the long run variance of  $a(L)^{-1}v_t$ , i.e.  $\omega^2 = \hat{\sigma}_v^2/a(1)^2$ .

In this setting, consider  $t_{\hat{\phi}}$  and  $t_{\hat{\gamma}}$  statistics, computed as in section 2.2 but on replacing (2.9) and (2.18) with the fitted regressions

$$\Delta\hat{u}_{y,t} = \hat{\phi}\hat{u}_{y,t-1} + \sum_{j=1}^p \hat{\pi}_j\hat{u}_{y,t-j} + \sum_{j=0}^q \hat{\delta}_j\hat{u}_{y,t-j} + \hat{\eta}_t, \quad (2.41)$$

$$\Delta\tilde{u}_{y,t} = \tilde{\phi}\tilde{u}_{y,t-1} + \sum_{j=1}^p \tilde{\pi}_j\tilde{u}_{y,t-j} + \sum_{j=0}^q \tilde{\delta}_j\tilde{u}_{y,t-j} + \tilde{\eta}_t \quad (2.42)$$

It can then be shown that the large sample results for  $t_{\hat{\phi}}$ ,  $t_{\hat{\gamma}}$  and  $UR(t_{\hat{\phi}}, t_{\hat{\gamma}})$  from sections 2.3 and 2.4 continue to hold. Moreover,  $R^2$  is now consistently estimated using the form of  $\hat{R}^2$  given in the previous section, but with  $\hat{e}_t$  and  $\hat{v}_t$  replaced with residuals from  $q$ 'th and  $p+1$ 'th order autoregressions fitted to  $u_{x,t}$  and  $u_{y,t}$ , respectively. In practice, since  $p$  and  $q$  are unknown, they can be determined endogenously using typical lag order selection rules such as downward testing or application of an information criterion.

## 2.6 Empirical illustration

In this section we provide two empirical examples to illustrate the behaviour of the Elliott and Jansson's (2003) and Hansen's (1995) unit root tests. The first empirical example tests the unit root null hypothesis on the U.S. industrial production series over the period 1973q1 – 2007q4. Industrial production is based on the Industrial Production Index which measures real output for manufacturing, mining and electric and gas utilities. The covariate used is the unemployment rate. The second empirical example performs the unit root tests on inflation calculated as the annualised growth rate of CPI i.e.  $\ln(CPI_t/CPI_{t-1}) \times 400$ , using the change of the long-term government bond yields as the stationary covariate, over the period 1969q3 – 2005q2. All data are obtained from the Federal Reserve Bank of St. Louis database.

### 2.6.1 Unit Root Test on Industrial Production

The first step in conducting this empirical illustration is to test the stationarity of covariate i.e. the unemployment series over the period 1973q1 – 2007q4. We employ the standard ADF test (not the covariate augmented ADF test), denoted by  $t_{ADF}$ . The lag selection was performed based on Ng and Perron (1995). The authors suggest setting an upper bound  $p_{\max}$  and estimating the ADF test regression with lag length,  $p$ , equal to  $p_{\max}$ . If the coefficient on the last lagged term is significant then the unit root test is performed with  $p = p_{\max}$ . Otherwise, the lag length is reduced by one and the process is repeated. Performing this procedure, we concluded that the ADF regression should be augmented by only one lag.

Moreover, since there is uncertainty about whether the series admits a trend or only a mean, we follow the procedure set by Enders (2004). As it is outlined in Ender’s (2004) chapter 4, p.213-214, we start with the least restrictive model in which both a trend and a mean are admitted i.e.

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \delta_1 \Delta x_{t-1} + \epsilon_t,$$

where  $x_t$  denotes the unemployment rate at time  $t$ , and use the ADF t-statistic,  $t_{ADF} = \frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$ , where  $\hat{\gamma}$  denotes the OLS estimate of  $\gamma$  and  $s.e.(\hat{\gamma})$  denotes the standard error of the coefficient, to test the unit root null hypothesis that  $\gamma = 0$ . Unit root tests based on less restrictive models which contain more deterministic components have low power to reject the null hypothesis; hence if the unit root null hypothesis is rejected, there is no need to determine whether too many deterministic regressors were included. If the unit root test rejects, the sequence  $\{x_t\}$  does not contain a unit root. In this example, the ADF t-statistic is equal to  $t_{ADF} = -3.97$  which is less than the 5% critical value of  $-3.45$  and therefore, we reject the unit root null hypothesis. Following Ender’s procedure, there is no need to examine whether too many deterministic regressors are included and we conclude that the unemployment rate is stationary over the period 1973q1 – 2007q4.

Individual covariate augmented unit root test that do not admit a trend in the industrial production series,  $y_t$ , do not reject the unit root null hypothesis across almost all start dates while rejections are often obtained for covariate augmented unit root tests that admit a trend, suggesting that a trend is present in the data. Besides, a mere observation of the industrial production index series clearly suggests a trend is present. As a result, inferences on the performance of unit root tests

based on tests that only admit a constant will be misleading given the misspecification. Moreover, such tests do not shed light on the sensitivity of the tests to the initial observation as the tests decisions do not really vary. In addition, we fit a trend in the covariate,  $x_t$ . We, therefore, illustrate the impact of the initial condition on the relative performance of test based on  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$ , under model model C, as summarised in section 2.2. To absorb serial correlation, we set  $p = 2$  and  $q = 2$  in equations (2.41) and (2.42) when computing the statistics  $t_{\hat{\phi}}$  and  $t_{\tilde{\phi}}$ .

We conduct the individual covariate augmented unit root tests and the size-corrected union of rejections strategy to the industrial production series repeatedly, each time moving the start date in order to observe the robustness of the test decisions to the initial condition. Specifically, we apply the OLS and GLS tests based on  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  and follow the union of rejections strategies, based on  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  procedures, in order to test for a unit root in the industrial production index using unemployment as a stationary covariate. To gauge the impact on the initial condition, we repeat the tests based on 50 consecutive start dates: 1973q1 – 1985q2 with common ending date 2007q4. The idea is that with different starting dates, we observe different values of the initial condition, thereby measuring the sensitivity of the tests to the size of the initial condition. Table 2.4 shows the tests results. The null rejections implied by these tests and procedures are reported in the last five columns of the table while the first column records the start date; the second column shows the estimated correlation coefficient for each regression and the third column shows the initial condition for each start date. To gauge the relevant size of the initial observation the following procedure is followed:

1) Regress the industrial production index (denoted by  $IP_t$ ) on a constant and a trend component:

$$IP_t = \mu_{IP} + \beta_{IP}t + \varepsilon_{IP,t}$$

2) Obtain the estimates,  $\hat{\mu}_{IP}$  and  $\hat{\beta}_{IP}$ , as well as the residuals,  $\hat{\varepsilon}_{IP,t}$ .

3) Obtain the standard error of the regression:

$$\hat{s}_{IP} = \sqrt{\frac{\sum_{t=1}^T \hat{\varepsilon}_{IP,t}^2}{T-2}}.$$

4) Compute the initial condition,  $IC$ , as:

$$IC = \frac{IP_1 - \hat{\mu}_{IP} - \hat{\beta}_{IP} * t}{\hat{s}_{IP}}$$

In order to easily observe the impact on initial condition on the tests decisions, the initial condition column in table 2.4 is sorted in ascending order. As it can be seen, the biggest initial conditions are reported in 1979 and early 1980, while the smallest initial conditions are reported for 1982 – 1983. The initial condition varies from 0.021 to 2.456, giving considerable scope to analyse the behaviour of the two tests over this range. Table 2.4 also shows that there is a strong correlation between the industrial production index and unemployment as  $\hat{R}^2$  is significantly large (above 0.5) in almost every case, indicating that significant power gains are achieved by including information from the covariate.

The significant gains achieved by the union of rejection strategy are pronounced as shown in table 2.4. The union of rejection strategy based on  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  consistently rejects the unit root null across the full range of start dates and only in 9 out of the 50 start dates considered fails to reject. Additionally, the procedure based on  $UR(t_{\hat{\phi}}, t_{\hat{\phi}})$  yields good results, rejecting the unit root null hypothesis in most cases. As explained in the previous sections, the  $t_{\hat{\phi}}$ –based test has low power across small values of the initial condition and the power is increasing as the magnitude of the initial condition increases, and results validate these findings to some extent. It is shown that the OLS test attains higher power than GLS tests for larger initial conditions, as the rejection frequency of the OLS test is much higher than that of GLS tests for large initial condition values. Results reported in table 2.4 also reflects the power properties of GLS-type tests. The  $t_{\hat{\phi}}$ –based and  $\hat{\Lambda}$ –based tests almost always reject for small values of the initial observation while they clearly fail to reject when the initial observation is big. The relative robustness of the union of rejection strategy, which rejects over the full range of the start dates, is clearly illustrated in this application.

Table 2.4: Covariate Augmented Unit Root Tests Outcomes on  
U.S. Industrial Production, using unemployment rate as covariate

Start Date	$\hat{R}^2$	Initial Condition	$t_{\hat{\phi}}$	$t_{\tilde{\phi}}$	$\hat{\Lambda}$	$UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$	$UR(t_{\hat{\phi}}, \hat{\Lambda})$
1983q1	0.973	0.021	R	R	-	R	R
1982q4	0.971	0.022	-	R	-	R	-
1983q2	0.979	0.147	R	R	-	R	R
1982q3	0.968	0.431	-	R	R	R	R
1975q2	0.943	0.441	-	R	R	R	R
1983q3	0.983	0.510	R	-	-	R	R
1975q3	0.946	0.587	-	-	R	-	R
1975q1	0.940	0.674	-	R	R	R	R
1975q4	0.945	0.704	-	R	R	R	R
1983q4	0.984	0.773	R	R	R	R	R
1982q2	0.969	0.804	-	R	-	R	-
1976q1	0.941	0.931	-	R	R	R	R
1985q2	0.984	0.976	R	R	R	R	R
1976q2	0.939	0.983	R	-	-	R	R
1976q3	0.944	1.039	R	R	R	R	R
1985:q1	0.983	1.091	R	R	R	R	R
1984:q1	0.984	1.123	R	R	R	R	R
1982:q1	0.972	1.136	-	R	R	R	R
1976:q4	0.948	1.158	R	R	R	R	R
1984:q4	0.983	1.168	R	R	R	R	R
1984:q2	0.983	1.263	R	R	R	R	R
1984:q3	0.984	1.273	R	R	R	R	R
1977:q1	0.946	1.305	R	R	-	R	R
1974:q4	0.906	1.454	-	-	-	-	-
1980:q3	0.969	1.481	R	-	-	R	R
1981:q4	0.973	1.575	-	R	R	R	R

Table 2.4 (Continued): Covariate Augmented Unit Root Tests Outcomes on  
U.S. Industrial Production, using unemployment rate as covariate

Start Date	$\hat{R}^2$	Initial Condition	$t_{\hat{\phi}}$	$t_{\tilde{\phi}}$	$\hat{\Lambda}$	$UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$	$UR(t_{\hat{\phi}}, \hat{\Lambda})$
1977q2	0.945	1.593	R	-	-	R	R
1978q1	0.935	1.614	R	-	-	R	R
1977q3	0.948	1.672	R	-	-	R	R
1977q4	0.947	1.701	R	-	-	R	R
1980q2	0.970	1.795	R	-	-	R	R
1973q1	0.863	1.886	R	-	-	R	R
1981q2	0.971	1.928	-	-	-	-	-
1974q3	0.855	1.935	-	-	-	-	-
1973q2	0.866	1.939	R	-	-	R	R
1981q1	0.971	1.941	R	-	-	R	R
1980q4	0.971	1.960	R	-	-	R	R
1973q3	0.856	2.003	R	-	-	R	R
1974q2	0.861	2.009	-	-	-	-	-
1981q3	0.972	2.010	-	-	-	-	-
1974q1	0.872	2.023	-	-	-	-	-
1978q2	0.937	2.048	R	-	-	R	R
1973:q4	0.867	2.139	R	-	-	R	R
1978:q3	0.936	2.140	R	-	-	R	R
1979:q3	0.939	2.358	R	-	-	R	R
1978:q4	0.941	2.362	R	-	-	R	R
1979:q4	0.939	2.403	-	-	-	-	-
1979:q2	0.938	2.404	R	-	-	R	R
1979:q1	0.940	2.423	R	-	-	R	R
1980:q1	0.955	2.456	R	-	-	R	R

### 2.6.2 Unit Root test on CPI Inflation

In the second empirical example, we test the unit root null hypothesis on the annualised CPI inflation rate using the change in the long-term government bond yields as covariate. As before, the first task is to perform a standard ADF unit root test on the government bond yields over the period 1969q3 – 2005q2. The lag selection procedure as outlined by Ng and Perron (1995) indicated a lag order of 7. Following the Enders (2004) procedure, we first perform a unit root null hypothesis for the extended model in which the ADF regression admits both a mean and a trend component:

$$\Delta g_t = \alpha + \beta t + \gamma g_{t-1} + \sum_{j=1}^7 \delta_j \Delta g_{t-j} + \epsilon_t,$$

where  $g_t$ , denotes the government bond yield at time  $t$ . Since the test statistic,  $t_{ADF} = -1.687$  is greater than the 5% critical value of  $-3.44$ , we do not reject the unit root null hypothesis. Since unit root tests which contain many deterministic components have low power, and given that the test failed to reject, we need to examine whether too many deterministic regressors are included. We, thus, perform a joint F-test to check the null hypothesis that  $\gamma = \beta = 0$ . Since the F-statistic, equal to 2.778, is less than the critical value of 3.064, we cannot reject the null hypothesis that the trend component is absent from the government bond yields series. As a result, we continue according to Enders and estimate a mean-only ADF regression:

$$\Delta g_t = \alpha + \gamma g_{t-1} + \sum_{j=1}^6 \delta_j \Delta g_{t-j} + \epsilon_t.$$

According to Ng and Perron (1995), the appropriate lag length in this case is 6. Since the ADF statistic on  $\gamma$  is  $t_{ADF} = 1.751$  which is greater than the 5% critical value of  $-2.89$ , we fail to reject the unit root null hypothesis and proceed with testing the joint hypothesis that  $\gamma = \alpha = 0$  using an F-test. The F-statistic is equal to 0.808, so we cannot reject that the series has zero-mean. We, therefore, estimate the ADF regression with 7 lags as indicated by Ng and Perron (1995) but without a mean or trend:

$$\Delta g_t = \gamma g_{t-1} + \sum_{j=1}^7 \delta_j \Delta g_{t-j} + \epsilon_t.$$

The ADF statistic is  $-1.272$  is not less than the 5% critical value of  $-1.95$  so we cannot reject the unit root null hypothesis. The Enders procedure stops here and we conclude that the government bond yields series has a unit root. To transform the series into a stationary one, we take the first

difference of the government bond yields series and perform a unit root test on the differenced series. Let  $\tilde{g}$  denote the differenced government bond yields series. We begin with the less restrictive model in which the ADF regression admits both a mean and a trend:

$$\Delta\tilde{g}_t = \alpha + \beta t + \gamma\tilde{g}_{t-1} + \sum_{j=1}^6 \delta_j \Delta\tilde{g}_{t-j} + \epsilon_t.$$

Since the ADF statistic is  $-5.911$  which is less than the 5% critical value of  $-3.44$ , we reject the unit root null hypothesis and conclude that the change in the government bond yields is a stationary series. If the least restrictive model, which has low power, was successful in rejecting the unit root null hypothesis, then we are confident that the sequence does not have a unit root.

Table 2.5 reports the results for the individual unit root tests and the size-corrected union of rejections strategy to the CPI inflation series for the mean case (model A in section 2.2). Again we perform the tests repeatedly each time moving the start date in order to observe the sensitivity of test outcomes to the initial condition. The initial condition here, shown in the 3<sup>rd</sup> column of table 2.5, is obtained by dividing the demeaned first observation by the standard error of all the observations in the sample for each regression. As it can be seen, the biggest initial conditions are reported in 1980 and 1981; a period with double digit inflation rates, while the smallest initial conditions are observed for the early 1970s.

We perform covariate augmented unit root tests and apply the union of rejections procedure on CPI inflation series using the change in government bond yields as covariate, based on 50 consecutive start dates: 1969q3 – 1981q4 and common ending point 2005q2. The superiority of the union of rejections strategies  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  is clearly demonstrated in table 2.5 since the strategies consistently reject the unit root null across the full range of the start dates and only in very few cases the procedures fail to reject. The GLS tests based on  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  always reject for small values of the initial condition whereas they rarely reject for large values of the initial condition with the OLS test based on  $t_{\hat{\phi}}$  behaving in exactly the opposite way. The union of rejection strategy, which maintains good power properties across the full range of initial condition values, gives a clear picture that individual tests fail to convey: The CPI inflation series is stationary.



Table 2.5: Covariate Augmented Unit Root Tests Outcomes on							
CPI Inflation using the change in government bond yields as covariate							
Start Date	$\hat{R}^2$	Initial Condition	$t_{\hat{\phi}}$	$t_{\hat{\phi}}$	$\hat{\Lambda}$	$UR(t_{\hat{\phi}}, t_{\hat{\phi}})$	$UR(t_{\hat{\phi}}, \hat{\Lambda})$
1976q1	0.209	0.103	-	R	R	R	R
1970q3	0.197	0.154	-	R	R	R	R
1975q2	0.225	0.159	-	R	R	R	R
1972q4	0.096	0.171	-	R	R	R	R
1971q3	0.146	0.207	-	R	R	R	R
1976q2	0.261	0.224	-	R	R	R	R
1969q3	0.081	0.264	-	R	R	R	R
1971q2	0.151	0.300	-	R	R	R	R
1970q2	0.169	0.300	-	R	R	R	R
1970q4	0.134	0.372	-	R	R	R	R
1971q1	0.089	0.400	-	R	R	R	R
1972q1	0.113	0.434	-	R	R	R	R
1977q3	0.125	0.436	-	R	R	R	R
1972q3	0.104	0.454	-	R	R	R	R
1969q4	0.082	0.472	-	R	R	R	R
1973q1	0.090	0.484	-	R	R	R	R
1976q4	0.159	0.507	-	R	-	R	-
1971q4	0.100	0.525	-	R	R	R	R
1977q4	0.122	0.551	-	R	R	R	R
1970q1	0.099	0.555	-	R	R	R	R
1972q2	0.111	0.644	-	R	R	R	R
1976q3	0.223	0.684	-	R	R	R	R
1978q1	0.143	0.868	-	R	R	R	R
1977q2	0.179	0.896	-	R	R	R	R
1973q3	0.121	0.979	R	R	R	R	R
1977q1	0.214	1.000	-	R	R	R	R

Table 2.5 (Continued): Covariate Augmented Unit Root Tests Outcomes on							
CPI Inflation using the change in government bond yields as covariate							
Start Date	$\hat{R}^2$	Initial Condition	$t_{\hat{\phi}}$	$t_{\hat{\phi}}$	$\hat{\Lambda}$	$UR(t_{\hat{\phi}}, t_{\hat{\phi}})$	$UR(t_{\hat{\phi}}, \hat{\Lambda})$
1975q4	0.273	1.004	-	-	-	-	-
1973q2	0.128	0.118	R	R	-	R	R
1975q3	0.287	1.205	-	R	R	R	R
1975q1	0.092	1.361	-	R	-	R	-
1978q2	0.129	1.585	-	-	-	-	-
1973q4	0.048	1.661	R	-	-	R	R
1978q3	0.107	1.679	R	-	-	R	R
1978q4	0.142	1.703	R	-	-	R	R
1980q3	0.002	1.881	R	-	-	R	R
1974q2	0.151	1.914	R	-	-	R	R
1979q1	0.160	1.987	R	-	-	R	R
1974q3	0.213	2.120	R	-	-	R	R
1981q4	0.137	2.150	R	-	-	R	R
1974q1	0.062	2.223	R	-	-	R	R
1974q4	0.111	2.497	-	-	-	-	-
1981q2	0.154	2.782	R	-	R	R	R
1979q2	0.027	2.907	R	-	-	R	R
1979q3	0.006	3.105	R	-	-	R	R
1979q4	0.343	3.194	R	-	-	R	R
1980q4	0.029	3.654	R	-	-	R	R
1981q1	0.092	3.884	R	-	-	R	R
1980q2	0.336	4.189	R	-	R	R	R
1981q3	0.807	4.483	R	-	-	R	R
1980q1	0.568	4.546	R	-	R	R	R

## 2.7 Conclusion

In this chapter we have considered the power of covariate augmented unit root tests, based on OLS demeaning/detrending and GLS demeaning/detrending, in the presence of asymptotically non-negligible initial conditions. In particular, based on a general structure for the initial condition, which encompasses the set-up of Müller and Elliott (2003) and Elliott and Müller (2006) as special cases, we find that where the initial condition is not asymptotically negligible, the quasi-differenced demeaned/detrended unit root tests of Elliott and Jansson (2003) can perform very badly indeed with their power against a given alternative rapidly decreasing towards zero as the size of the initial observation increases. Since we cannot be sure that such large initial conditions will not arise, this limits the reliability of such GLS-based tests in practice. In contrast, the OLS demeaned/detrended ADF tests of Hansen (1995) exhibit an increase in power as the size of the initial condition increases. This suggests that the Elliott and Jansson's tests are preferred when the initial condition is small while the Hansen's tests are far preferable when the initial condition is large. Consequently, a union of rejections decision rule between the Elliott and Jansson's tests and Hansen's tests could be fruitfully explored in order to maintain good power properties across both large and small initial conditions. Specifically, we followed the work of Harvey, Leybourne and Taylor (2009) whereby the unit root null hypothesis is rejected if either of the Elliott and Jansson's or Hansen test rejects. Reported asymptotic evidence suggests that our approach offers good robust power performance in the presence of uncertainty over the magnitude of the initial condition, retaining attractive power levels across zero, small and large initial condition magnitudes. Despite its impressive performance and efficiency, the strategy is easy to implement. Our findings mirror those found in the standard non-covariate augmented unit root testing environment, and our recommended procedure adds to the suite of available unit root testing procedures a covariate augmented approach that offers reliable power levels across the range of possible (unknown) initial conditions.

## 2.8 Appendix A

Here we provide a proof of the results for the most general model, Model C. The proofs for Models A and B follow in a similar fashion. In what follows we can set  $\mu_y = \mu_x = \beta_y = \beta_x = 0$  without

loss of generality, so that

$$\begin{aligned} y_t &= u_{y,t} \\ \Delta y_t &= \Delta u_{y,t} = cT^{-1}u_{y,t-1} + v_t \\ x_t &= u_{x,t} = e_t. \end{aligned}$$

Note that, in addition to the result in (2.21), we can also write

$$T^{-1/2}u_{y,\perp rT\perp} \Rightarrow \sigma_v K'_c(r)$$

where

$$K'_c(r) = \begin{cases} W_1(r) & c = 0 \\ \alpha e^{rc}/\sqrt{-2c} + W_{1c}(r) & c < 0 \end{cases}$$

i.e., for  $c < 0$ ,  $K'_c(r) = K_c(r) + \alpha/\sqrt{-2c}$ .

We first establish the part (i) result for  $t_{\hat{\phi}}$ . Here

$$\begin{aligned} \begin{bmatrix} T^{-1/2}\hat{\mu}_y \\ T^{1/2}\hat{\beta}_y \end{bmatrix} &= \begin{bmatrix} 1 & T^{-2}\sum_{t=1}^T t \\ T^{-2}\sum_{t=1}^T t & T^{-3}\sum_{t=1}^T t^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2}\sum_{t=1}^T u_{y,t} \\ T^{-5/2}\sum_{t=1}^T t u_{y,t} \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_v \int_0^1 K'_c(r) dr \\ \sigma_v \int_0^1 r K'_c(r) dr \end{bmatrix} \\ &= \begin{bmatrix} \sigma_v \left\{ 4 \int_0^1 K'_c(r) dr - 6 \int_0^1 r K'_c(r) dr \right\} \\ \sigma_v \left\{ 12 \int_0^1 r K'_c(r) dr - 6 \int_0^1 K'_c(r) dr \right\} \end{bmatrix} \end{aligned} \quad (2.43)$$

and

$$\begin{aligned} \begin{bmatrix} T^{1/2}\hat{\mu}_x \\ T^{3/2}\hat{\beta}_x \end{bmatrix} &= \begin{bmatrix} 1 & T^{-2}\sum_{t=1}^T t \\ T^{-2}\sum_{t=1}^T t & T^{-3}\sum_{t=1}^T t^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\sum_{t=1}^T e_t \\ T^{-3/2}\sum_{t=1}^T t e_t \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_e \left\{ RW_1(1) + \sqrt{1-R^2}W_2(1) \right\} \\ \sigma_e \int_0^1 r d \left\{ RW_1(r) + \sqrt{1-R^2}W_2(r) \right\} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_e P^C \\ \sigma_e Q^C \end{bmatrix}. \end{aligned} \quad (2.44)$$

Next,

$$\begin{aligned}\hat{\phi} &= \frac{\sum_{t=2}^T \hat{u}_{x,t}^2 \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{y,t-1} - \sum_{t=2}^T \Delta \hat{u}_{y,t-1} \hat{u}_{x,t} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{x,t}}{\sum_{t=2}^T \hat{u}_{y,t-1}^2 \sum_{t=2}^T \hat{u}_{x,t}^2 - \left( \sum_{t=2}^T \hat{u}_{y,t-1} \hat{u}_{x,t} \right)^2} \\ T\hat{\phi} &= \frac{T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{y,t-1} - T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t-1} \hat{u}_{x,t} T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{x,t}}{T^{-2} \sum_{t=2}^T \hat{u}_{y,t-1}^2 T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 - T^{-1} \left( T^{-1} \sum_{t=2}^T \hat{u}_{y,t-1} \hat{u}_{x,t} \right)^2}.\end{aligned}$$

Now consider each term in turn. First we find, given that  $\hat{\mu}_x = O_p(T^{-1/2})$  and  $\hat{\beta}_x = O_p(T^{-3/2})$  from (2.44),

$$\begin{aligned}T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 &= T^{-1} \sum_{t=2}^T (e_t - \hat{\mu}_x - \hat{\beta}_x t)^2 \\ &= T^{-1} \sum_{t=2}^T e_t^2 + o_p(1) \\ &\rightarrow \sigma_e^2.\end{aligned}\tag{2.45}$$

Next, using (2.43),

$$\begin{aligned}T^{-2} \sum_{t=2}^T \Delta \hat{u}_{y,t-1} &= T^{-2} \sum_{t=2}^T (u_{y,t} - \hat{\mu}_y - \hat{\beta}_y t)^2 \\ &= T^{-1} \sum_{t=2}^T \left\{ T^{-1/2} u_{y,t} - T^{-1/2} \hat{\mu}_y - T^{1/2} \hat{\beta}_y (T^{-1} t) \right\}^2 \\ &\Rightarrow \sigma_v^2 \int_0^1 L_c^C(r)^2 dr\end{aligned}\tag{2.46}$$

since it is straightforward to show that

$$\begin{aligned}&K'_c(r) - \left\{ 4 \int_0^1 K'_c(s) ds - 6 \int_0^1 s K'_c(s) ds \right\} - \left\{ 12 \int_0^1 s K'_c(s) ds - 6 \int_0^1 K'_c(s) ds \right\} r \\ &= K_c(r) - \left\{ 4 \int_0^1 K_c(s) ds - 6 \int_0^1 s K_c(s) ds \right\} - \left\{ 12 \int_0^1 s K_c(s) ds - 6 \int_0^1 K_c(s) ds \right\} r.\end{aligned}$$

We also find

$$\begin{aligned}
T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{y,t-1} &= T^{-1} \sum_{t=2}^T \Delta(u_{y,t} - \hat{\mu}_y - \hat{\beta}_y t)(u_{y,t-1} - \hat{\mu}_y - \hat{\beta}_y t) \\
&= T^{-1} \sum_{t=2}^T (u_{y,t-1} - \hat{\mu}_y - \hat{\beta}_y t) \Delta u_{y,t} \\
&\Rightarrow \sigma_v^2 \int_0^1 L_c^C(r) dK_c'(r)
\end{aligned} \tag{2.47}$$

$$\begin{aligned}
T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t-1} \hat{u}_{x,t} &= T^{-1} \sum_{t=2}^T (u_{y,t-1} - \hat{\mu}_y - \hat{\beta}_y t)(e_t - \hat{\mu}_x - \hat{\beta}_x t) \\
&\Rightarrow \sigma_e \sigma_v \int_0^1 L_c^C(r) d\{RW_1(r) + \sqrt{1-R^2}W_2(r)\} \\
&= \sigma_e \sigma_v R \int_0^1 L_c^C(r) dW_1(r) \\
&\quad + \sigma_e \sigma_v \sqrt{1-R^2} \int_0^1 L_c^C(r) dW_2(r)
\end{aligned} \tag{2.48}$$

and

$$\begin{aligned}
T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{x,t} &= T^{-1} \sum_{t=2}^T (cT^{-1}u_{y,t-1} + v_t - \hat{\beta}_y)(e_t - \hat{\mu}_x - \hat{\beta}_x t) \\
&= T^{-1} \sum_{t=2}^T v_t e_t + o_p(1) \\
&\rightarrow \sigma_{ev}
\end{aligned} \tag{2.49}$$

giving

$$\begin{aligned}
T\hat{\phi} &\Rightarrow \frac{\sigma_e^2 \sigma_v^2 \int_0^1 L_c^C(r) dK_c'(r) - \sigma_{ev} \sigma_e \sigma_v R \int_0^1 L_c^C(r) dW_1(r) - \sigma_{ev} \sigma_e \sigma_v \sqrt{1-R^2} \int_0^1 L_c^C(r) dW_2(r)}{\sigma_e^2 \sigma_v^2 \int_0^1 L_c^C(r)^2 dr} \\
&= c + (1+R^2) \frac{\int_0^1 L_c^C(r) dW_1(r)}{\int_0^1 L_c^C(r)^2 dr} - R\sqrt{1-R^2} \frac{\int_0^1 L_c^C(r) dW_2(r)}{\int_0^1 L_c^C(r)^2 dr}.
\end{aligned} \tag{2.50}$$

For s.e.  $(\hat{\phi})$ , we have

$$s.e.(\hat{\phi})^2 = \hat{\sigma}_\eta^2 \frac{\sum_{t=2}^T \hat{u}_{x,t}^2}{\sum_{t=2}^T \hat{u}_{y,t-1}^2 \sum_{t=2}^T \hat{u}_{x,t}^2 - (\sum_{t=2}^T \hat{u}_{y,t-1} \hat{u}_{x,t})^2}$$

$$\hat{\sigma}_\eta^2 = (T-1)^{-1} \sum_{t=2}^T \Delta(\hat{u}_{y,t} - \hat{\phi}\hat{u}_{y,t-1} - \hat{\delta}\hat{u}_{x,t})^2.$$

Consider the limit behaviour of  $\hat{\sigma}_\eta^2$ . First we have, using the results (2.45)-(2.49),

$$\begin{aligned} \hat{\delta} &= \frac{\sum_{t=2}^T \hat{u}_{y,t-1}^2 \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{x,t} - \sum_{t=2}^T \Delta \hat{u}_{y,t-1} \hat{u}_{x,t} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{y,t-1}}{\sum_{t=2}^T \hat{u}_{y,t-1}^2 \sum_{t=2}^T \hat{u}_{x,t}^2 - \left( \sum_{t=2}^T \hat{u}_{y,t-1} \hat{u}_{x,t} \right)^2} \\ &= \frac{T^{-2} \sum_{t=2}^T \hat{u}_{y,t-1}^2 T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{x,t} - T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t-1} \hat{u}_{x,t} T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{y,t-1}}{T^{-2} \sum_{t=2}^T \hat{u}_{y,t-1}^2 T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 - T^{-1} \left( \sum_{t=2}^T \hat{u}_{y,t-1} \hat{u}_{x,t} \right)^2} \\ &= \frac{T^{-1} \sum_{t=2}^T \Delta \hat{u}_{y,t} \hat{u}_{x,t}}{T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2} + o_p(1) \\ &\rightarrow \frac{\sigma_{ev}}{\sigma_e^2}. \end{aligned} \tag{2.51}$$

Then using

$$\begin{aligned} T^{-1} \sum_{t=2}^T (\Delta \hat{u}_{y,t})^2 &= T^{-1} \sum_{t=2}^T (\Delta u_{y,t} - \hat{\beta}_y)^2 \\ &= T^{-1} \sum_{t=2}^T (\Delta u_{y,t})^2 + o_p(1) \\ &= T^{-1} \sum_{t=2}^T v_t^2 + o_p(1) \\ &\rightarrow \sigma_v^2 \end{aligned}$$

along with (2.45)-(2.51), we find

$$\begin{aligned}
\hat{\sigma}_\eta^2 &= (T-1)^{-1} \sum_{t=2}^T \Delta(\hat{u}_{y,t} - \hat{\phi}\hat{u}_{y,t-1} - \hat{\delta}\hat{u}_{x,t})^2 \\
&= (T-1)^{-1} \sum_{t=2}^T (\Delta\hat{u}_{y,t})^2 + \hat{\delta}^2 (T-1)^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 + (T-1)^{-1} (T\hat{\phi})^2 T^{-2} \sum_{t=2}^T \hat{u}_{y,t-1}^2 \\
&\quad - 2(T-1)^{-1} T\hat{\phi}T^{-1} \sum_{t=2}^T \Delta\hat{u}_{y,t}\hat{u}_{y,t-1} - 2\hat{\delta}(T-1)^{-1} \sum_{t=2}^T \Delta\hat{u}_{y,t}\hat{u}_{x,t} \\
&\quad + 2(T-1)^{-1} \hat{\delta}T\hat{\phi}T^{-1} \sum_{t=2}^T \hat{u}_{y,t-1}\hat{u}_{x,t} \\
&= T^{-1} \sum_{t=2}^T (\Delta\hat{u}_{y,t})^2 + \hat{\delta}^2 T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 - 2\hat{\delta}T^{-1} \sum_{t=2}^T \Delta\hat{u}_{y,t}\hat{u}_{x,t} + o_p(1) \\
&\rightarrow \sigma_v^2 + \frac{\sigma_{ev}^2 \sigma_e^2}{\sigma_e^4} - 2 \frac{\sigma_{ev}^2}{\sigma_e^2} \\
&= \sigma_v^2 (1 - R^2)
\end{aligned}$$

and

$$\begin{aligned}
\left\{ T^{1/2} s.e.(\hat{\phi}) \right\}^2 &= \hat{\sigma}_\eta^2 \frac{T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2}{T^{-2} \sum_{t=2}^T \hat{u}_{y,t-1}^2 T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 - T^{-1} (T^{-1} \sum_{t=2}^T \hat{u}_{y,t-1} \hat{u}_{x,t})^2} \\
&\Rightarrow \frac{\sigma_v^2 (1 - R^2) \sigma_e^2}{\sigma_e^2 \sigma_v^2 \int_0^1 L_C^C(r)^2 dr} \\
&= \frac{1 - R^2}{\int_0^1 L_C^C(r)^2 dr}
\end{aligned}$$

giving the part (i) result

$$\begin{aligned}
t_{\hat{\phi}} &= \frac{T\hat{\phi}}{\sqrt{\left\{ T^{1/2} s.e.(\hat{\phi}) \right\}^2}} \\
&\Rightarrow \frac{c}{\sqrt{1 - R^2}} \sqrt{\int_0^1 L_c^i(r)^2 dr} + \sqrt{1 - R^2} \frac{\int_0^1 L_c^i(r) dW_1(r)}{\sqrt{\int_0^1 L_c^i(r)^2 dr}} - R \frac{\int_0^1 L_c^i(r)^2 dW_2(r)}{\sqrt{\int_0^1 L_c^i(r)^2 dr}}.
\end{aligned}$$

We now provide the proof of the part (ii) result for  $t_{\tilde{\phi}}$ . Here we have

$$\tilde{\phi} = \frac{\sum_{t=2}^T \hat{u}_{x,t}^2 \sum_{t=2}^T \Delta \tilde{u}_{y,t} \tilde{u}_{y,t-1} - \sum_{t=2}^T \Delta \tilde{u}_{y,t-1} \hat{u}_{x,t} \sum_{t=2}^T \Delta \tilde{u}_{y,t} \hat{u}_{x,t}}{\sum_{t=2}^T \tilde{u}_{y,t-1}^2 \sum_{t=2}^T \hat{u}_{x,t}^2 - \left( \sum_{t=2}^T \tilde{u}_{y,t-1} \hat{u}_{x,t} \right)^2}$$



$$T\tilde{\phi} = \frac{T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 T^{-1} \sum_{t=2}^T \Delta \tilde{u}_{y,t} \tilde{u}_{y,t-1} - T^{-1} \sum_{t=2}^T \Delta \tilde{u}_{y,t-1} \hat{u}_{x,t} T^{-1} \sum_{t=2}^T \Delta \tilde{u}_{y,t} \hat{u}_{x,t}}{T^{-2} \sum_{t=2}^T \tilde{u}_{y,t-1}^2 T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 - T^{-1} \left( T^{-1} \sum_{t=2}^T \tilde{u}_{y,t-1} \hat{u}_{x,t} \right)^2}.$$

In line with the results of Müller and Elliott (2003), it can be shown that

$$\begin{aligned} T^{-1/2} \tilde{u}_{y,\lfloor rT \rfloor} &\Rightarrow \sigma_v K_c(r) - \sigma_v \left\{ \bar{c}^* K_c(1) + 3(1 - \bar{c}^*) \int_0^1 s K_c(s) ds \right\} r \\ &= \sigma_v M_{c,\bar{c}}^C(r) \end{aligned}$$

from which we find

$$\begin{aligned} T^{-2} \sum_{t=2}^T \tilde{u}_{y,t-1}^2 &\Rightarrow \sigma_v^2 \int_0^1 M_{c,\bar{c}}^C(r)^2 dr \\ T^{-1} \sum_{t=2}^T \Delta \tilde{u}_{y,t} \tilde{u}_{y,t-1} &\Rightarrow \sigma_v^2 \int_0^1 M_{c,\bar{c}}^C(r) dM_{c,\bar{c}}^C(r) \\ &= \sigma_v^2 c \left[ \int_0^1 M_{c,\bar{c}}^C(r)^2 dr + ca/\sqrt{-2c} \int_0^1 M_{c,\bar{c}}^C(r) dr \right. \\ &\quad \left. + \int_0^1 M_{c,\bar{c}}^C(r) dW_1(r) + N_{c,\bar{c}}^C \left\{ c \int_0^1 r M_{c,\bar{c}}^C(r) dr - \int_0^1 M_{c,\bar{c}}^C(r) dr \right\} \right] \end{aligned}$$

where the last line is obtained using

$$\begin{aligned} dM_{c,\bar{c}}^C(r) &= dK_c(r) - N_{c,\bar{c}}^C dr \\ &= cK_c(r)dr + c\alpha\sqrt{-2c}dr + dW_1(r) - N_{c,\bar{c}}^C dr \\ &= cM_{c,\bar{c}}^C(r)dr + cN_{c,\bar{c}}^C r dr + c\alpha/\sqrt{-2c}dr + dW_1(r) - N_{c,\bar{c}}^C dr. \end{aligned}$$

The following results also obtain in a similar way as for OLS detrending:

$$\begin{aligned} T^{-1} \sum_{t=2}^T \tilde{u}_{x,t}^2 &\rightarrow \sigma_e^2 \\ T^{-1} \sum_{t=2}^T \Delta \tilde{u}_{y,t} \hat{u}_{x,t} &\rightarrow \sigma_{ev}. \end{aligned}$$

For the remaining term in  $T\tilde{\phi}$  we have

$$\begin{aligned}
T^{-1} \sum_{t=2}^T \tilde{u}_{y,t-1} \hat{u}_{x,t} &= T^{-1} \sum_{t=2}^T \tilde{u}_{y,t-1} (e_t - \hat{\mu}_x - \hat{\beta}_x t) \\
&= T^{-1} \sum_{t=2}^T \tilde{u}_{y,t-1} e_t - T^{1/2} \hat{\mu}_x T^{-3/2} \sum_{t=2}^T \tilde{u}_{y,t-1} - T^{3/2} \hat{\beta}_x T^{-5/2} \sum_{t=2}^T t \tilde{u}_{y,t-1} \\
&\Rightarrow \sigma_e \sigma_v \int_0^1 M_{c,\bar{c}}^C(r) d \left\{ RW_1(r) + \sqrt{1-R^2} W_2(r) \right\} \\
&\quad - \sigma_e \sigma_v R \int_0^1 M_{c,\bar{c}}^C(r) dW_1(r) + \sigma_e \sigma_v \sqrt{1-R^2} \int_0^1 M_{c,\bar{c}}^C(r) dW_2(r) \\
&\quad - \sigma_e \sigma_v P^C \int_0^1 M_{c,\bar{c}}^C(r) dr - \sigma_e \sigma_v Q^C \int_0^1 M_{c,\bar{c}}^C(r) dr.
\end{aligned}$$

So,

$$\begin{aligned}
&\sigma_e^2 \sigma_v^2 \left[ c \int_0^1 M_{c,\bar{c}}^C(r)^2 dr + c\alpha/\sqrt{-2c} \int_0^1 M_{c,\bar{c}}^C(r) dr + \int_0^1 M_{c,\bar{c}}^C(r) dW_1(r) \right. \\
&\quad \left. + N_{c,\bar{c}}^C \left\{ c \int_0^1 r M_{c,\bar{c}}^C(r) dr - \int_0^1 M_{c,\bar{c}}^C(r) dr \right\} \right] - \sigma_{ev} \sigma_e \sigma_v \left[ R \int_0^1 M_{c,\bar{c}}^C(r) dW_1(r) \right. \\
&\quad \left. + \sqrt{1-R^2} \int_0^1 M_{c,\bar{c}}^C(r) dW_2(r) - P^C \int_0^1 M_{c,\bar{c}}^C(r) dr - Q^C \int_0^1 r M_{c,\bar{c}}^C(r) dr \right] \\
T\tilde{\phi} &\Rightarrow \frac{\sigma_e^2 \sigma_v^2 \int_0^1 M_{c,\bar{c}}^C(r)^2 dr}{\sigma_e^2 \sigma_v^2 \int_0^1 M_{c,\bar{c}}^C(r)^2 dr} \\
&= c + (1-R^2) \frac{\int_0^1 M_{c,\bar{c}}^C(r) dW_1(r)}{\int_0^1 M_{c,\bar{c}}^C(r)^2 dr} - R\sqrt{1-R^2} \frac{\int_0^1 M_{c,\bar{c}}^C(r) dW_2(r)}{\int_0^1 M_{c,\bar{c}}^C(r)^2 dr} + \frac{c\alpha}{\sqrt{-2c}} \frac{\int_0^1 M_{c,\bar{c}}^C(r) dr}{\int_0^1 M_{c,\bar{c}}^C(r)^2 dr} \\
&\quad + \frac{N_{c,\bar{c}}^C \left\{ c \int_0^1 r M_{c,\bar{c}}^C(r) dr - \int_0^1 M_{c,\bar{c}}^C(r) dr \right\}}{\int_0^1 M_{c,\bar{c}}^C(r)^2 dr} + R \frac{P^C \int_0^1 M_{c,\bar{c}}^C(r) dr + Q^C \int_0^1 r M_{c,\bar{c}}^C(r) dr}{\int_0^1 M_{c,\bar{c}}^C(r)^2 dr}.
\end{aligned}$$

For  $s.e.(\tilde{\phi})$ , we again have

$$\begin{aligned}
\tilde{\delta} &\rightarrow \frac{\sigma_{ev}}{\sigma_e^2} \\
\tilde{\sigma}_v^2 &\rightarrow \sigma_v^2(1-R^2)
\end{aligned}$$

and hence

$$\begin{aligned}
T^{1/2}\{s.e.(\tilde{\phi})\}^2 &= \frac{\hat{\sigma}_\eta^2 \sum_{t=2}^T \hat{u}_{x,t}^2}{T^{-2} \sum_{t=2}^T \tilde{u}_{y,t-1}^2 T^{-1} \sum_{t=2}^T \hat{u}_{x,t}^2 - T^{-1} (T^{-1} \sum_{t=2}^T \tilde{u}_{y,t-1} \hat{u}_{x,t})^2} \\
&\Rightarrow \frac{\sigma_v^2 (1-R^2) \sigma_e^2}{\sigma_e^2 \sigma_v^2 \int_0^1 M_{c,\bar{c}}^C(r)^2 dr} \\
&= \frac{1-R^2}{\int_0^1 M_{c,\bar{c}}^C(r)^2 dr}
\end{aligned}$$

giving the part (ii) result

$$\begin{aligned}
t_{\tilde{\phi}} &= \frac{T\tilde{\phi}}{\sqrt{\{T^{1/2}s.e.(\tilde{\phi})\}^2}} \\
&\Rightarrow \frac{c}{\sqrt{1-R^2}} \sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr} + \sqrt{1-R^2} \frac{\int_0^1 M_{c,\bar{c}}^i(r) dW_1(r)}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} - R \frac{\int_0^1 M_{c,\bar{c}}^i(r)^2 dW_2(r)}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} \\
&\quad + \frac{c\alpha}{\sqrt{-2c(1-R^2)}} \frac{\int_0^1 M_{c,\bar{c}}^i(r) dr}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} + \frac{1}{\sqrt{1-R^2}} \frac{N_{c,\bar{c}}^i \{c \int_0^1 r M_{c,\bar{c}}^i(r) dr - \int_0^1 M_{c,\bar{c}}^i(r) dr\}}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} \\
&\quad + \frac{R}{\sqrt{1-R^2}} \frac{P^i \int_0^1 M_{c,\bar{c}}^i(r) dr + Q^i \int_0^1 M_{c,\bar{c}}^i(r) dr}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}}.
\end{aligned}$$

Finally, the part (iii) result for  $\hat{\Lambda}$  follows from the limit result in Theorem 1 of Elliott and Jansson (2003) for Case 5, on replacing  $W_{1c}(r)$  with  $K_c(r)$ .

## 2.9 References

1. Dickey, D.A. and W.A. Fuller (1979), "Distribution of the estimators for autoregressive time series with a unit root," Journal of the American Statistical Association, 74, pp.427-431.
2. Elliott, G. and M. Jansson (2003), "Testing for unit roots with stationary covariates," Journal of Econometrics, 115, pp.75-89.

3. Elliott, G. and U.K. Müller (2006), "Minimizing the impact of the initial condition on testing for unit roots," *Journal of Econometrics* 135, pp.285-310.
4. Elliott, G., T. J. Rothenberg and J. H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64 (4), pp.813-836.
5. Enders, W. (2004), "Applied Econometric Time Series," Second Edition, John Wiley & Sons: United States.
6. Hansen, B. E. (1995), "Rethinking the univariate approach to unit root testing using covariates to increase power," *Econometric Theory*, 11, pp.1148-1171.
7. Harvey, D.I, S.J. Leybourne and A.M.R. Taylor (2009), "Unit root testing in practice: Dealing with uncertainty over the trend and initial condition," *Econometric Theory*, 25, pp.587-636.
8. Müller, U.K. and G. Elliott (2003), "Tests for unit roots and the initial condition," *Econometrica*, 71, pp.1269-1286.
9. Ng, S. and P. Perron (1995), "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of American Statistical Association*, Vol. 90, Issue 429, pp.268-281.
10. Said, E. and D.A. Dickey (1984), "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order," *Biometrika*, 71, pp.599-607.
11. Stock, J.H. and M.W. Watson (1999), "Forecasting Inflation," *Journal of Monetary Economics*, 44, pp.293-335.
12. Westerlund, J. (2013), "A computationally convenient unit root test with covariates, conditional heteroskedasticity and efficient detrending," *Journal of Time Series Analysis*, 34, pp.477-495.

## Chapter 3

# Dealing with Uncertainty over both the existence of structural break and autocorrelation properties of the data

### 3.1 Introduction

In recent years a vast strand of literature is motivated by the apparent presence of breaks in the level and/or trend in macroeconomic time series (Stock and Watson (1996, 1999, 2005), Perron and Zhu (2005)). Correct specification of the break in the deterministic component of the series is extremely significant for modelling, estimation and forecasting. However, considerable structural uncertainty can be embedded when the existence of the break is not obvious or is questioned. This led to the development of a number of structural break tests (Andrews ((1993), Andrews and Ploberger (1994)) so that if the tests reject the no structural break hypothesis, a structural break model is estimated in the second stage. This two-step approach requires that a restricted estimator should be applied when the structural break test is insignificant, while the unrestricted estimator, that accounts for a break point, should be implemented once the test is significant. However, it has been shown that pretest estimators have poor sampling properties. In particular, the squared error of the pretest estimators is parameter-dependent and can be quite high relative to the unrestricted estimation (Hansen (2009)). Moreover, Harvey, Leybourne and Taylor (2012) show that in with-break unit root tests, that employ break detection methods based on auxiliary statistics (such as Carrion-i-Silvestre et al. (2009) and Harris et al. (2009)), valley power functions prevail in finite samples, despite the fact that these methods are near asymptotically efficient.

At the same time, uncertainty also exists as to whether the underlying stochastic component is best modelled by a stationary  $I(0)$  or unit root  $I(1)$  process. Vogelsang (1998), Harvey, Leybourne and Taylor (2009, 2010) and Perron and Yabu (2009) are just some of recent research directed to testing for structural breaks when the order of integration of the series is unknown. If the innovations are stationary, estimates are obtained by minimising the sum of squared residuals from a regression of the level of the series on the appropriate deterministic components. On the other hand, if it is known that a unit root is present, more efficient estimates are obtained by minimising the sum of squared residuals from a first-differenced version of the relevant regression (Harris et al. (2009)). Carrion-i-Silvestre, Kim and Perron (2009) also suggest minimising the sum of squared residuals based on a quasi-differenced regression when appropriate. Since the order of integration is usually unknown, one way to proceed is to pretest for a unit root and condition on the results of the pretests in performing second-stage break inference. However, as Elliott and Stock (1994) show in their Monte Carlo experiment, this procedure introduces substantial size distortions in the second-stage test. If the innovations of the regressor and the second-stage regression error are correlated, the size of the second stage test cannot be controlled effectively, even asymptotically. Moreover, performing a unit root test will not serve to answer what is the best filter that should be applied. It will only tell whether a unit root is present.

In this chapter, we consider estimating the level break size when the presence of a structural break is uncertain and when the order of integration of the data is unknown. The co-existence of these two types of uncertainties can severely undermine inferences on the behaviour of macroeconomic times series: Evidence in favour of unit roots can be a manifestation of structural changes and vice versa (Perron (1989)). In particular, the power of unit root tests is severely reduced when the process is stationary but subject to structural breaks (Perron (1989)), suggesting that a pre-test for the presence of a break is necessary, in which case all the undesirable sampling properties of pre-testing will have to be incurred. However, structural change tests will reject the no-change null hypothesis when the process contains a unit root with constant parameters, suggesting that a pre-test for the presence of unit roots is needed. Thus, a vicious circle is created and an apparent question arises: Should the researcher perform a unit-root test before testing for the presence of structural breaks or should he/she test for the presence of structural breaks before conducting unit

root tests?

In order to deal with this interplay and account for both of these kinds of uncertainties, we suggest a two-step procedure. The first stage involves the application of the practice developed by Harvey and Leybourne (2013) who select between first differences estimators and a number of quasi-differenced alternatives according to which achieves the smallest minimum sum of squared residuals. As it has been found, this approach achieves "most of the desirable properties of the appropriate estimators for the stationary and unit root worlds, without the inherent downsides involved in selecting purely one approach" (Harvey and Leybourne (2013, p.19)). While a unit-root test will only tell whether a unit root exists and will not point to the appropriate filtering of the data, this approach does not require a clear stand on the integration order, thus avoiding the need to defend a possible incorrect view on the series properties. Most importantly, the approach is conservative because a break is imposed a priori. If there is no break present in data, the approach is just inefficient. If on the other hand, there is a break, we do exactly what we should and thus avoid the detrimental effects of ignoring the break.

In the second stage, we follow work by Hansen (2009) and use a weighted average of the break size under a restricted model which does not allow for a break and, therefore, the break size is zero, and the unrestricted model which allows for a break and thus the break size is non-zero. Acknowledging the uncertainty over the existence of a structural break, Hansen (2009) avoids the inherent downsides of pre-testing and suggests a much better practice involving model averaging, with weights selected to minimise a modified Mallow information criterion (Mallow (1973)), constructed as an unbiased estimate of the in-sample fit and is a simple function of the sum of squared errors, the Andrews SupF test statistic and a penalty term.

The outline of the chapter is as follows: In section 3.2 we provide a literature review on the econometrics of structural break detection and estimation. In section 3.3 we consider the possibility of a single break in the level of the series and perform the two-step procedure, whereby in the first step we select the appropriate filtering parameter, while in the second step we use a weighted average of the estimators. The results from the Monte-Carlo experiments, which consider a range of autoregressive structures, are then compared to Hansen's approach that assumes that innovations are white noise. It is shown that the proposed estimator performs very well across a wide range

of possible DGPs, outperforming Hansen's unfiltered approach. Section 3.4 extends the approach to allow for two breaks in the level of the series. Section 3.5 of this chapter shows an empirical example which demonstrates the strength of the approach. Section 3.6 concludes. Due to the high volume of graphs in this chapter, all figures are shown at the end of the chapter in section 3.7. In what follows,  $[\cdot]$  denotes the integer part of its argument,  $\Rightarrow$  denotes weak convergence, and  $1(\cdot)$  denotes the indicator function.

## 3.2 Literature Review

### 3.2.1 Estimation and Inference about Break Dates

Bai (1997) and Bai and Perron (1998) develop a methodology explicitly designed for estimating and testing regression models with multiple breaks. Consider the multiple linear regression with  $m$  breaks:

$$y_t = x_t' \beta + z_t' \delta_j + u_t, \quad (3.1)$$

$t = T_{j-1}, \dots, T_j$ , for  $j = 1, \dots, m+1$  where  $x_t$  ( $p \times 1$ ) and  $z_t$  ( $q \times 1$ ) are vectors of covariates,  $\beta$  and  $\delta_j$  are the associated coefficients and  $u_t$  is the disturbance at  $t$ . The break points, denoted by  $(T_1, \dots, T_m)$  are treated as unknowns and the benchmark case is that  $T_0 = 0$  and  $T_{m+1} = T$ . As the authors emphasise, the aim is to estimate both the unknown regression coefficients and the break points, given the observed dataset  $(y_t, x_t, z_t)$ . This specification allows for both a partial structural change model, in which the parameter  $\beta$  is not subject to shifts and is therefore estimated based on the whole sample, and for a pure structural change model, where  $p = 0$  and all coefficients are subject to change. Using a partial structural change model can be beneficial in allowing potential savings in the number of degrees of freedom, which is particularly useful when we deal with multiple changes.

The multiple linear regression system (3.1) can be written in matrix form as:

$$Y = X\beta + \bar{Z}\delta + U,$$

where  $Y = (y_1, \dots, y_T)'$ ,  $X = (x_1, \dots, x_T)'$ ,  $U = (u_1, \dots, u_T)'$ ,  $\delta = (\delta_1', \delta_2', \dots, \delta_{m+1}')'$  and  $\bar{Z}$  is the matrix which diagonally partitions  $Z$  at  $(T_1, \dots, T_m)$ . Let the true value of the parameter be denoted with



a 0 superscript so that the true data-generating process is given by:

$$Y = X\beta^0 + \bar{Z}^0\delta^0 + U \quad (3.2)$$

Equation (3.2) is estimated using least squares. For each  $m$  partition  $(T_1, \dots, T_m)$ , the least squares estimates of  $\beta$  and  $\delta_j$  are generated by minimising the sum of squared residuals, i.e.

$$S_T(T_1, \dots, T_m) = (Y - X\beta - \bar{Z}\delta)'(Y - X\beta - \bar{Z}\delta) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - x'_t - z'_t\delta_i]^2. \quad (3.3)$$

Let  $\hat{\beta}(\{T_j\})$  and  $\hat{\delta}(\{T_j\})$  denote the estimates based on the given  $m$ -partition  $(T_1, \dots, T_m)$ , denoted by  $\{T_j\}$ . Substituting these in the objective function, equation (3.3), the estimated breakpoints are generated by:

$$(\hat{T}_1, \dots, \hat{T}_m) = \arg \min_{(T_1, \dots, T_m)} S_T(T_1, \dots, T_m).$$

Thus, the breakpoint estimators correspond to the global minimum of the sum of squared residuals objective function. Having estimated the breakpoints, the corresponding least squares parameters are calculated as  $\hat{\beta}(\{\hat{T}_j\})$  and  $\hat{\delta}(\{\hat{T}_j\})$ .

### Restrictions on the parameters

Perron and Qu (2006) consider multiple structural changes allowing for linear restriction on the parameters of the conditional mean. Consider the model:

$$y = \bar{Z}\delta + u,$$

where

$$R\delta = r,$$

with  $R$  a  $k \times (m+1)q$  matrix with rank  $k$  and  $r$ , a  $k$  dimensional vector of constants. A partial structural model can arise as a special case when restricting some coefficients to remain unchanged across regimes. This provides some flexibility when estimating models which specifies a number of states less than the number of regimes, with some coefficients remaining stable in consecutive regimes. Perron and Qu (2006) show that the same consistency and rate of convergence results prevail while the limit distribution of the break point estimates remain unaffected by the imposition of parameter restrictions. An important advantage of their method is that when valid restrictions are imposed, the tests become more powerful.

## Dynamic Programming and Global Minimisers

Estimation of structural breaks models requires global minimisers of the objective function (3.3) with least squares operations of order  $O(T^m)$  which is difficult for multiple structural breaks. Bai and Perron (2003) developed an efficient algorithm to obtain global minimisers of the sum of squared residuals based on dynamic programming requiring least squares operations of order  $O(T^2)$  at most for any number of breaks.

The first step is to compute the sum of squared residuals of the relevant segments and use the dynamic algorithm to detect the partition which yields a global minimisation of the overall sum of squared residuals. The method requires a sequential detection of the optimal one-break partitions. Let  $SSR(\{T_{r,n}\})$  denote the sum of squared residual obtained from the optimal partition containing  $r$  breaks using the first  $n$  observations. The optimal partition solves the recursive problem:

$$SSR(\{T_m, T\}) = \min_{mh \leq j \leq T-h} [SSR(\{T_{m-1,j}\}) + SSR(J+1, T)]$$

The optimal one-break partition is first evaluated for all sub-samples that allow a possible break ranging from observations  $h$  to  $T - mh$ , storing the optimal one-break partitions and associated sum of squared residuals. Each partition corresponds to sub-samples ending at dates ranging from  $2h$  to  $T - (m-1)h$ . The next step is to evaluate optimal partitions with two breaks corresponding to sub-samples ending at dates ranging from  $3h$  to  $T - (m-2)h$ . The method searches for the one-break partition minimising the sum of squared residuals for each ending date giving a set of optimal two-breaks partitions. The procedure continues until a set of optimal  $(m-1)$  breaks partitions are obtained with ending dates ranging from  $(m-1)h$  to  $T - 2h$ . The final step is to see which of the optimal  $(m-1)$  break partitions yields an overall minimal sum of squared residuals when combined with an additional segment. For models with restrictions and partial structural change models, the dynamic programming method to obtain global minimisers of the sum of squared residuals cannot be applied directly, so a different method is followed based on a simple iterative procedure.

## The limit distribution of break dates estimates

Bai (1997) and Bai and Perron (1998, 2003) show that a number of factors affect the limit distributions of the break dates estimates which include:

- (1) The size of shift in the coefficients, with larger shifts increasing accuracy.
- (2) The sample moment matrices of the regressors for the segments before and after the true break date.
- (3) The long-run variance of  $\{w_t u_t\}$ , which allows for serial correlation in the errors
- (4) The existence of trends in the regressor series.

Confidence intervals are then constructed based on the consistently estimated nuisance parameters.

### **Estimating Breaks sequentially**

Bai (1997) and Bai and Perron (1998) show that when allowing for only one break model when in fact the regression experiences multiple breaks, the estimated break fraction converges to the true break fraction which allows for the greatest reduction in the sum of squared residuals. The second one break model can then be estimated, after imposing the first identified break, so that it converges to the second most dominant true break point, the one which allows for the second greatest reduction in the sum of squared residuals. In particular, Bai and Perron (1998) explain that when the number of breaks points is known, for example  $m$ , and the first break point is identified, the sample is split into two sub-samples separated by the estimated break point. For each sub-sample, a one break is estimated and the second breakpoint is chosen which allows the greatest reduction in the sum of squared residuals. The sample is then split into three regimes and one-break model is estimated in each regime. The third break point is selected such that the reduction in the sum of squared residuals is the biggest. The process continues until the  $m$  breaks are identified.

Bai (1997) also shows that the limit distribution of the estimates obtained sequentially differs from the limit distribution derived from the simultaneous estimation of the break points. For example, when break dates are estimated simultaneously, the limit distribution of the date depends exclusively on the parameters of the regimes next to it, while when breaks are estimated sequentially, the limit distributions depend on the parameters of all regimes. Re-estimating each break date conditional on the adjacent break dates, as suggested by Bai (1997), makes the limit distributions under both estimation methods identical.

### 3.2.2 Tests for structural change

There is a long history of diagnostic testing based on fitted econometric models to see if the fitted parameters are stable across the available sample data. Here we mention the main ones that are more closely related to our work.

#### Tests for parameter constancy in models that do not allow for a break

**The CUSUM test** The recursive least squares estimates of  $\beta$  are based on estimating

$$y_t = \beta'_t x_t + \xi_t, \quad t = 1, \dots, n$$

by least squares recursively for  $t = k + 1, \dots, n$  giving  $n - k$  least squares estimates  $(\hat{\beta}_{k+1}, \dots, \hat{\beta}_T)$ .

The recursive least squares estimates can be efficiently computed using the Kalman filter. If there is no structural change,  $\beta$  is constant over time and all  $n - k$  estimates will have a common value.

Let the recursive residuals be noted by:

$$\begin{aligned} w_t &= \frac{v_t}{\sqrt{f_t}} = \frac{y_t - \beta'_{t-1} x_t}{\sqrt{f_t}} \\ f_t &= \sigma^2 [1 + x'_t (X'_t X_t)^{-1} x_t], \end{aligned}$$

where  $X_t$  contains the observations on the regressors up to time  $t$ . Then the CUSUM statistic of Brown, Durbin and Evans (1975) is given by:

$$CUSUM_t = \sum_{j=k+1}^t \frac{\hat{w}_j}{\hat{\sigma}_w},$$

where  $\hat{\sigma}_w^2 = \frac{1}{n-k} \sum_{t=1}^n (w_t - \bar{w})^2$ . Under the null hypothesis that  $\beta$  is constant, CUSUM statistic has mean zero and variance that is proportional to  $t - k - 1$ . The limit distribution of the CUSUM test can be expressed in terms of the maximum of a weighted Wiener process i.e.

$$CUSUM \rightarrow \sup_{0 \leq t \leq 1} \left| \frac{W(t)}{1 + 2t} \right|,$$

where  $W(t)$  is a unit Wiener process defined on  $(0, 1)$ .

**The CUSUMsq test** Brown, Durbin and Evans (1975) provide an alternative, the CUSUM of squares test whose statistic is given by:

$$CUSUMSQ_t = \frac{\sum_{j=k+1}^t \hat{w}_j^2}{\sum_{j=k+1}^n \hat{w}_j^2}.$$

Ploberger and Krämer (1990) considered the local power function of  $CUSUMQ_t$  and show that it has power equal to size for local changes that specify a one-time change in the coefficients.

**Generalised Fluctuation tests** These tests were developed by Ploberger, Krämer and Kontrus (1989) and Kuan and Hornik (1995). They examine the maximum difference between the OLS estimate of  $\beta$  using the full sample and the OLS estimates using subsets of the sample, either chosen recursively or by a moving window that rolls through the sample at constant width. If the true coefficients remain constant over time i.e. there is no structural break, the two types of estimates should be similar. The resulting processes should not fluctuate too much under the null. If, on the contrary, the process exhibits large fluctuations, there is evidence of structural break.

A problem with the fluctuation tests as with the CUSUMSQ test is that for a given sample size, the power function are not monotonic and might decrease as the alternative becomes further away from the null value. For example, considering a one-off shift in the mean, it has been shown that the power of the tests decreases as the shift becomes larger and might even reach zero. This can be attributed to different scaling of the variance under the null and the alternative hypotheses. As Perron (2006) explains, because the break is not modelled, the variance estimated is contaminated by the shift under the alternative and as the magnitude of the shift becomes greater, the estimate gets inflated and, therefore, we experience losses in power. Overall, among the tests considered, CUSUM is to be preferred since it avoids these drawbacks.

**The Nyblom test** Consider the stochastic time-varying parameter models (TVP) in which we treat the breaks in  $\beta$  as stochastic or the result of a continuous process. The regression equation can then be written as

$$\begin{aligned} y_t &= \beta_t' X_t + e_t, \quad t = 1, \dots, T \\ \beta_t &= \beta_{t-1} + v_t, \end{aligned} \tag{3.4}$$

where  $\beta_0 = \beta$  and where  $e_t$  and  $v_t$  are uncorrelated and  $v_t$  is serially uncorrelated with  $E(v_t'v_t) = \tau^2 G$ . Under the null hypothesis,  $\beta$  is constant i.e.  $H_0 : E(v_t'v_t) = 0$  for all  $t$  while under the alternative  $E(v_t'v_t) > 0$  for some  $t$ . In the context of (3.4) our alternative of interest is that  $H_1 : v_t \sim IIDN(0, \tau^2 G)$ , where  $G$  is a known  $k$  by  $k$  matrix. Here  $\beta_t$  evolves as a random walk, so it evolves smoothly but randomly over the sample period. It can be shown that the Lagrange multiplier statistic of  $H_0$  against  $H_1$  in (3.4) is given by:

$$L := T^{-1} \sum_{t=1}^T S_T(s/T)' \left( \frac{\hat{\sigma}^2}{T} \sum_{t=1}^T X_t X_t' \right)^{-1} S_T(s/T),$$

where  $\hat{\sigma}^2 := T^{-1} \sum \hat{e}_t^2$  and  $S_T(\lambda) := T^{-1/2} \sum_{R=[T\lambda]+1}^T \hat{e}_t X_t$ , and where  $\{\hat{e}_t\}_{t=1}^T$  are OLS residuals from regressing  $y_t$  on  $X_t$ ,  $t = 1, \dots, T$ . The decision rule is to reject the null hypothesis if the  $L$  statistic is large and greater than the critical value. The above test is for the choice of  $G = [T^{-1} \sum_{t=1}^T X_t X_t']^{-1}$  which makes for a neat limiting distribution. Critical values are calculated by simulation by Nyblom (1989). It is important to note that the approach only works if we assume that  $X_t$  is stationary. If there is a deterministic trend or the regressors contain unit roots, different distribution of  $L$  applies. Moreover, the test is applicable for models estimated by methods other than OLS. A drawback of this model is that the test is not informative about the date or the type of structural change.

### Tests allowing for a single break in the model

**The Chow (1960) test** Suppose we have the classical linear regression model

$$y_t = \beta_t' X_t + e_t, \quad t = 1, \dots, T \quad (3.5)$$

and we wish to test the null  $H_0 : \beta_t = \beta$  against the alternative  $H_1 : \beta_t = \beta, t \leq r$  and  $\beta_t + \gamma, t > r$ . Classically, when  $\{X_t, e_t\}$  satisfy standard stationarity and regularity conditions we can use the Chow (1960) test. Assuming for simplicity that  $e_t \sim IID(0, \sigma^2)$  and  $X_t$  is fixed, the Chow test is given by:

$$F_t(r/T) := \frac{SSR_{1,T} - (SSR_{1,r} + SSR_{r+1,T})}{(SSR_{1,r} + SSR_{r+1,T})/(T - 2k)},$$

where  $SSR_{a,b}$  is the residual sum of squares from estimating (3.5) on the sample observations  $a, \dots, b$  and  $k$  is the dimension of  $\beta_t$ . For fixed  $r/T$ , i.e. for a known break point,  $F_t(r/T) \xrightarrow{w} x_k^2$  under  $H_0$ . We reject the null hypothesis of no structural break if the sum of squared residuals corresponding to

the model with no breaks is significantly greater than the sum of squared residuals corresponding to the model with a break at period  $r$ . The problem with the Chow (1960) test is that is not operational if the breakpoint is unknown i.e. when we do not hold strong beliefs about the exact timing of the break.

**The Quandt (1960) test** When the break point is unknown, i.e.  $r$  is not known, the problem becomes non-standard since as Perron (2006) explains one parameter is only identified under the alternative hypothesis because under the null, there is no break. This is known as the Davies problem (1977). Quandt (1960) introduces the sup F test and suggests estimating (3.5) over a range of possible dates  $r_0, \dots, r_1$ , and maximise over these, giving a likelihood ratio test. In other words, Quandt (1960) formulates a likelihood ratio test for a change in parameters evaluated at the break date that maximises the likelihood function. Thus, the test statistic is given by:

$$QLR := \max_{r=r_0, \dots, r_1} \{F_T(r/T)\}.$$

Davies (1977) showed that if estimated parameters are unidentified under the null, standard  $\chi^2$  inference does not obtain. Kim and Siegmund (1989), therefore, found that the limiting distribution of the QLR under the  $H_0$  is given by:

$$QLR \xrightarrow{w} \sup_{\lambda \in [\lambda_0, \lambda_1]} \left\{ \frac{B_k^\mu(\lambda)' B_k^\mu(\lambda)}{\lambda(1-\lambda)} \right\},$$

where  $B_k^\mu(r) := W_k(r) - rW_k(1)$ , where  $W_k(r)$  is a  $k$ -dimensional Brownian motion and  $B_k^\mu(r)$  is a  $k$ -dimensional Brownian bridge. Also,  $\lambda_i := \lim_{T \rightarrow \infty} r_i/T$ ,  $i = 0, 1$ . The trimming parameters  $\lambda_0$  and  $\lambda_1$  must be set so that we cannot have  $\lambda_0 = 1$  and  $\lambda_1 = 1$  because breaks are hard to identify near the beginning and end of the sample. Andrews (1993) suggests  $\lambda_0 = 0.15$  and  $\lambda_1 = 0.85$  if there is no knowledge of the break date. Implicitly, the break date  $r$  and the break fraction  $\lambda$  are estimated using

$$\begin{aligned} \hat{r} &= \arg \max_r F_T(r/T) \\ \hat{\lambda} &= \hat{r}/T. \end{aligned}$$

Andrews (1993) shows that if the trimming parameters  $\lambda_0 = \lambda_1 = 0$  so that no restrictions are imposed, the test diverges to infinity under the null hypothesis indicating that critical values increase

while the power of the test decreases as the trimming parameters get smaller. As Perron (2006, p.25) note, "the range over which we search for a maximum must be small enough for the critical values not to be too large and for the test to retain descent power, yet large enough to include break dates that are potential candidates." Critical values for a range of dimensions and for trimming intervals are tabulated by Andrews (1993).

**Andrews (1993) test** Andrews (1993) also considers tests based on the maximal value of the Wald and LM tests showing that they have the same limit distribution under the null and under a number of alternative hypotheses. Tests considered are consistent and have non trivial local asymptotic power against alternatives for which the parameters of interest exhibit instability over the interval specified by  $\lambda_0$  and  $\lambda_1$ .

### Optimal tests

**Andrews and Ploberger (1994) test** Optimal tests which maximise the weighted average power have been designed by Andrews and Ploberger (1994). Two weighting schemes have been applied. The one applies weights to the parameter identified under the alternative assigning a weight function  $J(\lambda_1)$  which can be viewed as a prior distribution over the possible break dates. The other scheme considers the distance between the alternative value and the null hypothesis in an asymptotic framework which treats the alternative values as local to the null. The optimal test is a weighted function of the test statistics for all permissible fixed break dates. In particular, optimal tests turn out to be weighted averages of the Chow breakpoint statistics  $F_T(\frac{r}{T})$  used to compute the QLR statistic:

$$\begin{aligned} ExpF_T &= \ln\left(\frac{1}{r_2 - r_1 + 1}\right) \sum_{t=r_1}^{r_2} \exp\left(\frac{1}{2}k \cdot F_T\left(\frac{t}{T}\right)\right) \\ AveF_t &= \frac{1}{r_2 - r_1 + 1} \sum_{t=r_1}^{r_2} k \cdot F_T\left(\frac{t}{T}\right), \end{aligned}$$

where  $k$  is the number of regressors being tested. Asymptotic null distributions are non-standard and depend on  $k$ ,  $\lambda_0$  and  $\lambda_1$ . The critical values are tabulated in Andrews and Ploberger (1994) for symmetric trimmings. In general using either of the basic statistics (i.e. Wald, LM, LR) leads



to tests that are asymptotically equivalent. Simulations reported by Andrews, Lee and Ploberger (1996) show that the tests perform well in practice.

### 3.3 One Level-Break

#### 3.3.1 Model and Estimation

In this section, the model to be estimated is a linear time-series regression with a *possible* structural break in the level. The model for estimation is:

$$\begin{aligned} y_t &= \alpha_1 + \gamma_1 DU_t(\tau^*) + u_t, \quad t = 1, \dots, T, \\ u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 2, \dots, T, \end{aligned} \tag{3.6}$$

with  $u_1 = \varepsilon_1$ , where  $DU_t(\tau^*) = 1(t > \lfloor \tau^* T \rfloor)$  with  $\lfloor \tau^* T \rfloor$  the break point with associated break fraction  $\tau^*$  and level break magnitude  $\gamma_1$ . Here,  $\tau^*$  is unknown but satisfies  $\tau^* \in \Lambda$ , where  $\Lambda = [\tau_L, \tau_U]$  with  $0 < \tau_L < \tau_U < 1$ . We assume that the innovation process  $\{\varepsilon_t\}$  of equation (3.7) is an IID sequence with variance  $\omega_\varepsilon^2$  and finite fourth moment.

Uncertainty over the existence (and subsequently the location) of the break means that the researcher does not know whether the break-size is different from zero. If there is no break in the level of the series, then  $\gamma_1 = 0$ , and the model simplifies to:

$$\begin{aligned} y_t &= \alpha_1 + u_t, \quad t = 1, \dots, T \\ u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 2, \dots, T. \end{aligned} \tag{3.7}$$

With pre-testing for the existence of a break, the researcher had to make a firm decision between model (3.6) and model (3.7). However, pre-testing is shown to exhibit bad sample-properties (Hansen (2009)) and, in this context, is not recommended. At the same time, the researcher is uncertain about the autocorrelation structure of the series, and thus about the correct filtering parameter,  $\rho$ . In order to deal with both uncertainty over the existence of the break and uncertainty over the autocorrelation properties of the data, we employ a two-step procedure in which the first step involves a conservative method which allows for a break and selects the appropriate filtering parameter in order to absorb serial correlation. The second step involves estimating both the

model with a break, as characterised by (3.6) and the model without a break (model (3.7)) so that a weighted average of the break size is obtained, (Hansen (2009)). In what follows we describe the procedure in full.

### Step 1: Selecting the appropriate filtering parameter

The approach is conservative so despite the fact that the structural break is uncertain, the first step involves selecting the filtering parameter based on the equation that allows a structural break. This means that if the true data generating process does not involve a break, the approach is just inefficient. If, on the other hand, the true process involves a break, the approach achieves the best results and avoids the detrimental effects of excluding the break a priori. Since the true value of  $\rho$  is unknown in practice, we use Harvey and Leybourne's (2013) hybrid fraction estimator that selects between a range of possible values for  $\rho$  in the set  $D_m = \{\bar{\rho}_1, \bar{\rho}_2, \dots, \bar{\rho}_{m-1}, 1\}$ . The estimate of the break fraction  $\tau^*$  is then selected by minimising the residual sum of squares from a quasi-differenced version of equation (3.6), that is

$$\hat{\tau}_{D_m} = \arg \min_{\tau \in \Lambda, \bar{\rho} \in D_m} S(\bar{\rho}, \tau)$$

where  $S(\bar{\rho}, \tau)$  denotes the residual sum of squares from an OLS regression of  $\mathbf{y}_{\bar{\rho}}$  on  $\mathbf{Z}_{\bar{\rho}, \tau}$  with

$$\mathbf{y}_{\bar{\rho}} = [y_1, y_2 - \bar{\rho}y_1, \dots, y_T - \bar{\rho}y_{T-1}]',$$

$$\mathbf{Z}_{\bar{\rho}, \tau} = [\mathbf{x}_1, \mathbf{x}_2 - \bar{\rho}\mathbf{x}_1, \dots, \mathbf{x}_T - \bar{\rho}\mathbf{x}_{T-1}]' \text{ with } \mathbf{x}_t = [1, DU_t(\tau)]'.$$

More explicitly, the procedure simultaneously estimates the break fraction and selects a filtering parameter according to which achieves the smallest residual sum of squares. As Harvey and Leybourne (2013) show, the procedure works extremely well in selecting a  $\bar{\rho}$  which is almost identical to the true  $\rho$ , even in sample sizes of practical relevance. The hybrid estimator is shown to perform competitively against the better of the levels- and first differenced-based estimators across a range of I(0) and I(1) data generating processes since, as the authors explain, the autoregressive filtering inherent in  $\hat{\tau}_{D_m}$  is only intended to remove the dominant autoregressive behaviour present in a series and not to whiten the series entirely.

## Step 2: Weighted Average Break Size

Having obtained the relevant information from the first step, we proceed with dealing with uncertainty over the existence of the break. Based on the filtered data, if we believe a structural break has occurred then we estimate equation (3.6) by least squares written as:

$$\mathbf{y}_{\bar{p}} = \mathbf{Z}_{\bar{p},\tau} \hat{\beta}(\hat{\tau}) + \hat{u}_t(\hat{\tau}) \quad (3.8)$$

where

$$\hat{\beta}(\hat{\tau}) = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\gamma}_1 \end{bmatrix}. \quad (3.9)$$

$\hat{u}_t(\hat{\tau}) = \hat{\mathbf{u}}$  then denotes the  $T \times 1$  residuals from equation (3.8) while the sum of squared errors, given the estimated break fraction,  $\hat{\tau}$ , is given by  $\hat{u}_t(\hat{\tau})' \hat{u}_t(\hat{\tau})$ . To estimate the break fraction,  $\tau$ , based on equation (3.8) one needs to minimise the criterion:

$$\hat{\tau} = \arg \min_{\tau_L \leq \tau \leq \tau_U} \hat{u}_t(\hat{\tau})' \hat{u}_t(\hat{\tau}).$$

It is important to note here that the estimated break fraction,  $\hat{\tau}$ , is exactly equal to  $\hat{\tau}_{D_m}$  (the first stage estimated break fraction) because this second stage estimate minimises the sum of squared residuals conditional on the filtering parameter. Since the selected filtering parameter corresponds to the global minimum sum of squared residuals (across the whole range of filtering parameters considered), it is intuitive that  $\hat{\tau} = \hat{\tau}_{D_m}$ .

On the other hand, if we are certain that no structural break has occurred then we define

$$\bar{\mathbf{x}}_t = [1]' \text{ and } \bar{\mathbf{Z}}_{\bar{p},\tau} = [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2 - \bar{\rho}\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_T - \bar{\rho}\bar{\mathbf{x}}_{T-1}]'$$

and estimate equation (3.7) by least squares written as

$$\mathbf{y}_{\bar{p}} = \bar{\mathbf{Z}}_{\bar{p},\tau} \tilde{\beta} + \tilde{u}_t \quad (3.10)$$

where

$$\tilde{\beta} = \begin{bmatrix} \tilde{\alpha}_1 \end{bmatrix} \quad (3.11)$$

and  $\tilde{u}_t = \tilde{\mathbf{u}}$  now represents the  $T \times 1$  residuals from equation (3.10). Since no break is allowed in the level, the break size,  $\tilde{\gamma}_1$ , is equal to zero.

The standard practice in testing model (3.6) against model (3.7) is based on Andrews (1993) supF test. The test statistic is given by

$$F = \frac{(\tilde{\mathbf{u}}'\tilde{\mathbf{u}} - \hat{\mathbf{u}}'\hat{\mathbf{u}})}{s^2} \quad (3.12)$$

where

$$s^2 = \frac{1}{T-2} \hat{\mathbf{u}}'\hat{\mathbf{u}} \quad (3.13)$$

is the bias-corrected estimator of the error variance from model (3.8). This model selection procedure dictates that for an  $\alpha\%$  significance level, the test rejects model (3.7) in favour of model (3.6), if  $F \geq c_\alpha$ , where  $c_\alpha$  is the  $(1 - \alpha)\%$  upper quantile of the distribution of  $F$ . In other words, the pretest estimator uses the unrestricted estimator (3.9) when  $F$  is significant and otherwise uses the restricted estimator (3.11). So a pretest estimator for the level-break size can be written as:

$$\begin{aligned} \hat{\gamma}_1^p &= \hat{\gamma}_1 1(F \geq c_\alpha) + 0 \cdot 1(F < c_\alpha) \\ &= \hat{\gamma}_1 1(F \geq c_\alpha). \end{aligned}$$

Hansen (2009) has shown that averaging based on Mallows weights exhibits substantial efficiency gains over selection between the two models based on pretesting. In particular, the asymptotic mean-squared error of the pretest estimator is very large, for certain regions of the parameter space, while the asymptotic mean-squared error of their proposed weighted average estimator is close to the infeasible minimum. Averaging assigns a weight  $w$  to model (3.6), the structural break model, and a weight  $1 - w$  to model (3.7). Hansen (2009) suggests minimising a variant of the Mallows criterion given by:

$$C(w) = (\hat{\mathbf{u}}w + \tilde{\mathbf{u}}(1 - w))'(\hat{\mathbf{u}}w + \tilde{\mathbf{u}}(1 - w)) + 2s^2(1 + \bar{p}w) \quad (3.14)$$

where  $\bar{p}$  is a feasible penalty coefficient constructed as an average of limiting cases and depends on the number of structural breaks allowed and  $\lambda = \frac{(1 - \tau_L)^2}{\tau_L^2}$ . Hansen tabulates the penalty coefficients as function of the number of structural breaks and the trimming parameter  $\tau_L$ . The Mallows weight is the value in  $[0, 1]$  that minimises  $C(w)$ . Hansen (2009) demonstrates that optimal weights are given by:

$$\hat{w} = \begin{cases} 0 & \text{if } F < \bar{p} \\ 1 - \frac{\bar{p}}{F} & \text{if } F \geq \bar{p}. \end{cases} \quad (3.15)$$

Our analysis is based on Hansen's weighted average estimates of the model parameters as weighted averages using the weight  $\hat{w}$ . The level-break size is, then, given by:

$$\hat{\gamma}_1^{Hyb} = \hat{\gamma}_1(1 - \frac{\bar{p}}{F})1(F \geq \bar{p}).$$

### 3.3.2 Finite sample performance

In this section we compare the finite sample performance of the hybrid estimators  $\hat{\gamma}_1^{Hyb}$ , which is based on the filtered data, with the estimator suggested by Hansen based on unfiltered data. In particular, Hansen assumes that innovations are white noise so that the weighted average estimators of the break size are based on unfiltered data. In what follows, we show results for a range of autocorrelation structures, namely, for  $\rho = \{0, 0.5, 0.9, 1\}$ . As we will subsequently show, the results of the two procedures, i.e. Hansen's and our proposed hybrid procedure, are identical when  $\rho = 0$ . This is natural given that our approach only differs to Hansen's when innovations are not white. The simulation DGP is equation (3.6), with  $u_1 = \varepsilon_1$ ,  $\varepsilon_t \sim NIID(0, 1)$  and  $\alpha_1 = 0$  (without loss of generality) while we use  $\tau^* = 0.5$ . In the simulations here and in the remainder of the chapter, we set  $\Lambda = [0.15, 0.85]$ . We consider break sizes in the set  $\gamma_1 = \{1, 1.5, 2, -2.5\}$ . Regarding  $D_m$ , we set  $D_m = \{0, 0.1, 0.2, \dots, 0.8, 0.9, 0.95, 0.98, 0.99, 1\}$ . To perform the Hansen approach, we use the penalty term value of 2.49, as suggested in Hansen (2009, p.1506). The results are based on 1,000 simulations and sample size of  $T = 200$ . All simulations were programmed in Gauss 9.0.

Figures 3.1-3.14 compare histograms of the estimates of both the break fraction and the level break size under the approach of Hansen (2009) to the corresponding histograms developed under our new proposed hybrid approach, for different data-generating processes. Figures 3.1 and 3.2 represent histograms for the  $I(0)$  case for  $\gamma_1 = 1$  and  $\gamma_1 = -2.5$  respectively. What is immediately apparent is the fact that the histograms of the two approaches are identical. The reason is obvious: Because the approach by Harvey and Leybourne (2013) is extremely effective in capturing the true autoregressive parameter, our hybrid approach applied the Hansen's method based on unfiltered data. This means that the two approaches deliver the same results. What we further observe is

that as the break magnitude becomes larger, the more effective are the methods in picking up the correct break-fraction. However, the accuracy of estimating the correct break-size does not improve with the actual size of the break. The two approaches systematically detect the true break-size with probability of 62%.

The relative performance of the two approaches can be observed as soon as we allow some serial correlation. Figures 3.3-3.6 represent the histograms under the two approaches for  $\rho = 0.5$ . As we can see, the hybrid approach is always better than Hansen's approach in estimating the break-fraction and the superiority becomes stronger as the break-size becomes larger. As far as the break-size estimates are concerned, the mean estimates of both approaches are very close to the true break-size while the variance does not seem to differ much. However, as the true-break size becomes larger, the hybrid approach detects the true-break size with greater accuracy than Hansen's. For instance, in figure 3.8 in which the true-break size is  $-2.5$ , the hybrid approach detects the true-break size with probability 56% compared to Hansen's approach which accurately estimates the break-size with 50% probability.

As mentioned before, as the autoregressive behaviour becomes more dominant, i.e. as  $\rho$  becomes larger, our approach dramatically does better than Hansen's unfiltered approach in accurately estimating the break size. This is expected since the stronger autoregressive component begins to erode Hansen's estimator ability to identify the true break size. This becomes even more apparent in histograms represented in figures 3.7-3.10, which concern the case of  $\rho = 0.9$ , so that the series is nearly integrated of order one. As an indication, consider figure 3.7 which represents results for a break size of 1. The Hansen's approach is completely wrong in estimating the true break-fraction, as it implies that the break fraction is more likely to have occurred at break fraction 0.8, compared to 0.5 which is the truth. The hybrid approach is more accurate in predicting the break-fraction, which explains why it does better in estimating the break-size. As the break-size becomes bigger, the advantages of the hybrid approach are further highlighted. Figure 3.10, which shows results for  $\gamma_1 = -2.5$ , demonstrates that the hybrid approach correctly picks up the break-size 26% of the times, compared to Hansen's approach which detects the true break-size only 18% of the times.

Results are similar but more pronounced when we consider the I(1) case i.e. when  $\rho = 1$ , represented by histograms in figures 3.11-3.14. As it is expected, the superiority of our approach is

further highlighted. In particular, the majority of break size estimates based on the hybrid approach is much closer to the true break size than the majority of estimates suggested by Hansen's procedure. For instance, consider figure 3.13b which represents results for a level break-size of 2. The hybrid procedure is much more consistent in capturing the break-size compared to Hansen's procedure which delivers a very big variance. More concretely, the Hansen procedure fails to identify whether the break is negative or positive. On the contrary, the hybrid estimates correctly detect the true break-size with probability 20%. Even though this probability is not high enough to draw firm conclusions about the true-break size, the other estimates of the break-size lie within a very close range from the true break-size of 2. This superiority is likely to be stemmed from the excellent accuracy advantages of the hybrid approach in estimating the break-fraction. As we see from figures 3.11-3.14, while Hansen's break-fraction estimates are all over the place ranging from 0.15 to 0.85 with almost equal probability (see figure 3.14a for the most obvious example), the hybrid approach captures the true break-fraction with a far greater accuracy.

### **Allowing a simultaneous break in level and trend**

The analysis so far considered a level-break in the series. Here, we show that the two-stage hybrid procedure does not work equally well when we allow for a broken trend. In particular, the hybrid approach does not deliver significant improvements in the accuracy of trend break-size estimation. To show this, we consider a model that allows a simultaneous break in the level and trend:

$$y_t = \alpha_1 + \alpha_2 t + \gamma_1 DU_t(\tau^*) + \gamma_2 DT_t(\tau^*) + u_t, \quad t = 1, \dots, T,$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 2, \dots, T$$

with  $u_1 = \varepsilon_1$ , where  $DU_t(\tau^*) = 1(t > \lfloor \tau^* T \rfloor)$  and  $DT_t(\tau^*) = 1(\lfloor \tau^* T \rfloor)(t - \lfloor \tau^* T \rfloor)$  with  $\lfloor \tau^* T \rfloor$  the break point with associated break fraction  $\tau^*$  and level and trend break magnitudes  $\gamma_1$  and  $\gamma_2$  respectively. Here,  $\tau^*$  is unknown but satisfies  $\tau^* \in \Lambda$ , where  $\Lambda = [\tau_L, \tau_U]$  with  $0 < \tau_L < \tau_U < 1$  as before. If there is no break in the level and trend, then  $\gamma_1 = \gamma_2 = 0$ , and the model simplifies to

$$y_t = \alpha_1 + \alpha_2 t + u_t, \quad t = 1, \dots, T$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 2, \dots, T.$$

We follow the two-step procedure as outlined above: We first select the autoregressive filter parameter that minimises the sum of squared residuals and then average the break size estimators according to Hansen's (2009) weight scheme. Figure 3.15 shows the results for the I(1) case setting  $\gamma_1 = 3$  and  $\gamma_2 = 3$  and  $\tau^* = 0.5$ . As it can be seen, our hybrid approach detects the break-fraction with great accuracy: 98% of the times, the estimates of the hybrid approach are exactly equal to the true break-fraction, outperforming Hansen's unfiltered approach which accurately estimates the true break-fraction only 48% of the times. Due to its superiority in estimating the break-fraction, the hybrid approach clearly outperforms Hansen's methodology in estimating the *level*-break size. The Hansen approach delivers level break-size estimates that range from  $-15$  to  $10$  occurring at equal frequencies while the observation with the highest frequency is  $-1$ , well below the true level-break size of  $3$ . On the contrary, our hybrid approach yields a much smaller variance for the estimates while the mean of the estimates is close to  $3$ . More explicitly, the mean level break-size estimate given by Hansen's approach is  $-0.45$  with variance  $7.19$  while our hybrid approach yields a mean level-break size of  $2.70$  with variance  $1.66$ . Despite the outstanding accuracy in estimating the level-break size, the hybrid approach does not seem to perform any better in estimating the trend-break size. The mean is  $2.93$  for both approaches and the variance is very similar. This indicates that in a univariate framework, such as the one examined in this chapter, the hybrid approach is not worth pursuing when there is a broken trend.

### 3.4 Two Level-Breaks

Our proposed approach can be extended to consider multiple structural breaks in the levels. Although analysis in Hansen (2009) only considers breaks that happen simultaneously, we show that the approach also works well when we multiple consecutive breaks are allowed. In what follows we explain how the hybrid procedure works with two structural breaks in the level.



### 3.4.1 Model and Estimation

In this section, the model to be estimated is a linear time-series regression with two *possible* structural breaks in the level. The model for estimation is:

$$\begin{aligned} y_t &= \alpha_1 + \gamma_1 DU_{1t}(\tau_1^*) + \gamma_2 DU_{2t}(\tau_2^*) + u_t, \quad t = 1, \dots, T, \\ u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 2, \dots, T, \end{aligned} \quad (3.16)$$

with  $u_1 = \varepsilon_1$ , where  $DU_{1t}(\tau_1^*) = 1(t > \lfloor \tau_1^* T \rfloor)$  and  $DU_{2t}(\tau_2^*) = 1(t > \lfloor \tau_2^* T \rfloor)$  with  $\lfloor \tau_1^* T \rfloor$  and  $\lfloor \tau_2^* T \rfloor$  the break points with associated break fractions  $\tau_1^*$  and  $\tau_2^*$  and level break magnitudes  $\gamma_1$  and  $\gamma_2$ . Here,  $\tau_1^*$  and  $\tau_2^*$  are unknown but satisfy  $\tau_1^*, \tau_2^* \in \Lambda$ , where  $\Lambda = [\tau_L, \tau_U]$  with  $0 < \tau_L < \tau_U < 1$ . We assume that the innovation process  $\{\varepsilon_t\}$  of equation (3.16) is an IID sequence with variance  $\omega_\varepsilon^2$  and finite fourth moment.

Uncertainty over the existence of the breaks means that the researcher does not know whether the break-sizes are different from zero. If there are no breaks in the level of the series, then  $\gamma_1 = \gamma_2 = 0$ , and the model simplifies to

$$\begin{aligned} y_t &= \alpha_1 + u_t, \quad t = 1, \dots, T \\ u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 2, \dots, T. \end{aligned} \quad (3.17)$$

In order to deal with both uncertainty over the existence of the break and uncertainty over the autocorrelation properties of the data, we extend the two-step procedure outlined in the previous section to allow for two-level breaks. The procedure is summarised below.

#### Step 1: Selecting the appropriate filtering parameter

The approach is conservative so despite the fact that the structural breaks are not definite, the first step involves selecting the filtering parameter based on the equation that allows for two structural breaks in the level of the series. This means that if the true data generating process does not involve breaks, the approach is just inefficient. If, on the other hand, the true process involves level-breaks, the approach achieves the best results and avoids the detrimental effects of ignoring structural breaks a priori. Since the true value of  $\rho$  is unknown in practice, we use Harvey and Leybourne's (2013) hybrid fraction estimator that selects between a range of possible values for  $\rho$

in the set  $D_m = \{\rho'_1, \rho'_2, \dots, \rho'_{m-1}, 1\}$  where  $|\rho'_i| < 1$  for all  $i$  and, without loss of generality,  $-1 < \rho'_1 < \rho'_2, \dots, \rho'_{m-1} < 1$ . The estimates of the break fractions  $\tau_1^*$  and  $\tau_2^*$  given by  $\hat{\tau}_{D_m} = \{\hat{\tau}_{1D_m}, \hat{\tau}_{2D_m}\}$  are then selected by minimising the residual sum of squares from a quasi-differenced version of equation (3.16), that is

$$\hat{\tau}_{D_m} = \arg \min_{\tau \in \Lambda, \bar{\rho} \in D_m} S(\bar{\rho}, \tau),$$

where  $S(\bar{\rho}, \tau)$  denotes the residual sum of squares from an OLS regression of  $\mathbf{y}_{\bar{\rho}}$  on  $\mathbf{Z}_{\bar{\rho}, \tau}$  with

$$\mathbf{y}_{\bar{\rho}} = [y_1, y_2 - \bar{\rho}y_1, \dots, y_T - \bar{\rho}y_{T-1}]',$$

$$\mathbf{Z}_{\bar{\rho}, \tau} = [\mathbf{x}_1, \mathbf{x}_2 - \bar{\rho}\mathbf{x}_1, \dots, \mathbf{x}_T - \bar{\rho}\mathbf{x}_{T-1}]' \text{ and } \mathbf{x}_t = [1, DU_{1t}(\tau), DU_{2t}(\tau)]'.$$

The procedure simultaneously estimates the break fractions and selects a filtering parameter according to which achieves the smallest residual sum of squares. As we subsequently show, the procedure delivers extremely good results in selecting an autoregressive parameter which is very close to the true  $\rho$  even in small sample sizes across a range of I(0) and I(1) data generating processes.

## Step 2: Weighted Average Break Size

Once we obtain the filtered data based on information prevailed in the first step, we move on to deal with uncertainty over the existence of the break. Based on the filtered data, if we believe that structural breaks have taken place in the level, we estimate equation (3.16) by least squares written as:

$$\mathbf{y}_{\bar{\rho}} = \mathbf{Z}_{\bar{\rho}, \tau} \hat{\beta}(\hat{\tau}) + \hat{u}_t(\hat{\tau}) \quad (3.18)$$

where

$$\hat{\beta}(\hat{\tau}) = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix}. \quad (3.19)$$

$\hat{u}_t(\hat{\tau}) = \hat{\mathbf{u}}$  then denotes the  $T \times 1$  residuals from equation (3.18) while the sum of squared errors, given the estimated break fractions,  $\hat{\tau} = \{\hat{\tau}_1, \hat{\tau}_2\}$  are given by  $\hat{u}_t(\hat{\tau})' \hat{u}_t(\hat{\tau})$ . To estimate the break

fractions based on equation (3.16) one needs to minimise the criterion:

$$\hat{\tau} = \arg \min_{\tau_L \leq \hat{\tau} \leq \tau_U} \hat{u}_t(\hat{\tau})' \hat{u}_t(\hat{\tau}).$$

It is important to note here that the estimated break fractions,  $\hat{\tau}$ , are exactly equal to  $\hat{\tau}_{D_m}$  (the first stage estimated break fractions) because this second stage estimate minimises the sum of squared residuals conditional on the filtering parameter. Since the selected filtering parameter corresponds to the global minimum sum of squared residuals (across the whole range of filtering parameters considered), it is intuitive that  $\hat{\tau} = \hat{\tau}_{D_m}$ .

On the other hand, if no breaks are assumed, the regressors are as follows:

$$\bar{\mathbf{x}}_t = [1]' \quad \text{and} \quad \bar{\mathbf{Z}}_{\bar{\rho}, \tau} = [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2 - \bar{\rho}\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_T - \bar{\rho}\bar{\mathbf{x}}_{T-1}]'$$

and estimate equation (3.17) based on the filtered data:

$$\mathbf{y}_{\bar{\rho}} = \bar{\mathbf{Z}}_{\bar{\rho}, \tau} \tilde{\beta} + \tilde{u}_t, \quad (3.20)$$

where

$$\tilde{\beta} = \begin{bmatrix} \tilde{\alpha}_1 \end{bmatrix} \quad (3.21)$$

and  $\tilde{u}_t = \tilde{\mathbf{u}}$  now represents the  $T \times 1$  residuals from equation (3.20). Since no breaks are allowed, the level break sizes,  $\tilde{\gamma}_1 = \tilde{\gamma}_2 = 0$ , are equal to zero.

The standard practice in testing model (3.16) against model (3.17) is based on Andrews (1993) supF test. The test statistic is given by equations (3.12) and (3.13). The pretest estimator uses the unrestricted estimator (3.19) when F is significant and otherwise uses the restricted estimator (3.21). The pretest estimators for the two level-break sizes can then be written as:

$$\begin{aligned} \hat{\gamma}_1^p &= \hat{\gamma}_1 1(F \geq c_\alpha) + 0 \cdot 1(F < c_\alpha) \\ &= \hat{\gamma}_1 1(F \geq c_\alpha) \end{aligned}$$

$$\begin{aligned} \hat{\gamma}_2^p &= \hat{\gamma}_2 1(F \geq c_\alpha) + 0 \cdot 1(F < c_\alpha) \\ &= \hat{\gamma}_2 1(F \geq c_\alpha). \end{aligned}$$

Although Hansen (2009) does not show the analytics for two consecutive structural breaks, the approach is shown to perform competitively when extended to allow for two structural breaks in

the level. Instead of selecting purely between model (3.16) and (3.17), we follow Hansen (2009) and apply the model averaging technique where weights are selected to minimise the Mallows criterion (3.14). The weights are given by (3.15) where  $\bar{p}$  is the penalty coefficient corresponding to the two-break case. The weighted average level-break sizes are, then, given by:

$$\begin{aligned}\hat{\gamma}_1^{Hyb} &= \hat{\gamma}_1(1 - \frac{\bar{p}}{F})1(F \geq \bar{p}) \\ \hat{\gamma}_2^{Hyb} &= \hat{\gamma}_2(1 - \frac{\bar{p}}{F})1(F \geq \bar{p}).\end{aligned}$$

### 3.4.2 Finite sample performance

In this section we compare the finite sample performance of the hybrid estimates  $\hat{\gamma}_1^{Hyb}$  and  $\hat{\gamma}_2^{Hyb}$  with Hansen's estimates which are based on unfiltered data. In what follows, we show results for a range of autocorrelation structures, namely, for  $\rho = \{0, 0.5, 0.9, 1\}$ . The simulated DGP is equation (3.16), with  $u_1 = \varepsilon_1$ ,  $\varepsilon_t \sim NIID(0, 1)$  and  $\alpha_1 = 0$  (without loss of generality) while we use  $\tau_1^* = 0.25$  and  $\tau_2^* = 0.75$ . We consider  $\gamma_1 = \{3, 4, -5, 6\}$  and  $\gamma_2 = \{3, 4, 5, -6\}$ . As before, we set  $D_m = \{0, 0.1, 0.2, \dots, 0.8, 0.9, 0.95, 0.98, 0.99, 1\}$ . The results are based on 1,000 simulations and sample size of  $T = 200$ . Although Hansen (2009) does not tabulate penalty terms for the two consecutive breaks, we use 4.05 which is the penalty term for two simultaneous breaks, acknowledging that the penalty term is not likely to differ much.

Figures 3.16 and 3.17 show histograms for  $\rho = 0$ . As it can be seen, the results for the two procedures are almost identical since our proposed approach only differs to Hansen's approach when innovations are not white. Since data require no filtering when  $\rho = 0$ , and given that the hybrid approach correctly selects  $\bar{p} = 0$ , the two procedures yield identical results. As with the one break case, the two approaches work well in identifying the break-fractions. Indeed, the estimated break-fraction is equal to the true break-fraction more than 95% of the occasions.

Figures 3.18-3.21 show results for the  $\rho = 0.5$  case. Although it is not immediately apparent, the hybrid approach performs marginally better in estimating the two break fractions and, therefore, in estimating the break-sizes. However, because the degree of autocorrelation is not strong enough to erode the ability of Hansen's approach in detecting the true break-fractions, the difference between the two approaches is not dramatic. The advantages of the hybrid approach become more apparent

in figures 3.22-3.25, which present results for  $\rho = 0.9$ . The hybrid approach is much more effective in detecting the true break fraction and this is why estimates of the break-sizes are much more accurate. For instance, consider figure 3.25b, where the true break-sizes are 6 and  $-6$  for the first and second break respectively. The hybrid approach correctly estimates the first break-size of 6 with probability 22% compared to Hansen's approach which accurately estimates the first break-size with probability 14%. In addition, the hybrid procedure is right in estimating the second break-size 20% of the times compared to Hansen's procedure which correctly estimates the second break-size with probability of 14%. Although the difference is not huge, the variance and range of the break-size estimates under the hybrid approach are substantially smaller compared to Hansen's procedure.

The superiority of the hybrid approach becomes even more acute when we consider the I(1) case (figures 3.26-3.29). Results are substantial: The Hansen procedure completely fails to identify the location of the two breaks. On the contrary, the hybrid approach, which applies the correct filter of the data, is particularly effective in identifying the true break-fractions at 0.25 and 0.75. The difference in performance between the two approaches becomes even more apparent when we consider larger break sizes. For instance, consider figure 3.29 where the first level break-size is 6 and the second level break-size is  $-6$ . Even though the mean estimate of the first break-size under the Hansen procedure is 5.20, which is close to the true break size, the variance is 7.89. Indeed, Hansen's methodology correctly estimates the first break-size only 4% of the times. On the contrary, the hybrid approach delivers accurate results 20% of the times. Hansen's estimates of the second break size are all over the place, ranging from  $-25$  to  $20$ . By contrast, the corresponding histogram for the hybrid approach is tall and thin, without extreme values and estimates concentrating around  $-6$ .

### 3.5 Empirical Illustration

In this section, we provide an empirical example to illustrate the ability of the proposed hybrid approach to detect breaks and evaluate their sizes when the autocorrelation properties of the data are unknown and when the existence of breaks is uncertain. We employ quarterly data on the

seasonally adjusted U.S. civilian unemployment rate over the period 1965 $q1$  to 2014 $q2$ , obtained from the Federal Reserve Bank of St. Louis database. Since the seminal paper by Blanchard and Summers (1986) that introduced the hypothesis of unemployment hysteresis, according to which cyclical fluctuations will have a permanent effect on the level of unemployment, a number of studies have been conducted to examine the autocorrelation properties of the unemployment series. Evidence however is mixed: Among others, Elmskov and MacFarlan (1993); Mitchell (1993), and Camarero and Tamarit (2004) all conclude that unemployment rates are non-stationary while authors like Papell et al. (2000) and Camarero et al. (2006) provide evidence that unemployment is mean-reverting for the majority of countries examined. Could this disagreement be attributed to the fact that many studies ignore the possibility of structural breaks? Even though numerous studies have attempted to locate structural breaks in the unemployment rate of a particular country, there is no consensus about the existence of the break at a particular point in time, suggesting that considerable uncertainty also prevails about whether a break has taken place. Oil crises, changes in monetary policy and central bank operating procedures, important policy changes, such as trade agreements and tax reforms, are just some of potential sources of structural breaks in the unemployment series. Despite that, current procedures in identifying breaks do not lead to the same conclusions about the location of breaks.

We analyse the U.S. unemployment rate over the period 1965 $q1$  to 2014 $q2$ , allowing for two consecutive breaks in the level of the series. Figure 3.30 plots the U.S. civilian unemployment rate over the period considered. A mere observation indicates that the series only admits a level and that a trend is absent<sup>1</sup>. We follow the procedure outlined in previous section, using the trimming parameters  $\Lambda = [0.15, 0.85]$ , so that we only search for breaks over the period 1972 $q2$ –2007 $q1$ . Step 1 from the hybrid approach selects the autoregressive parameter to be equal to 0.98, indicating that the unemployment rate is nearly integrated, and it is, therefore, not surprising that unit root tests against stationary alternatives have low power (see also Cochrane (1991) and DeJong et al. (1992)). The results of the hybrid approach also point that the first structural break in the level occurred in 1974 $q4$ , while the second structural break in the level occurred in 1983 $q3$ . These findings are

---

<sup>1</sup>Clemente et al. (2005) and Lee and Chang (2008) provide evidence that the unemployment series admits only a level.

consistent with a number of studies that provided evidence of a mean shift in the unemployment series around the reported dates (Gil-Alana (2002), Arestis et al. (1999) Clemente et al. (2005), Lee et al. (2009)). The average break-sizes, as found by the hybrid procedure, are 1.53 for the first break and  $-0.76$  for the second break. The first structural break occurred around the first energy crisis that caused a significant and deep recession. As it can be seen from figure 3.30, in the new regime, starting from 1974q4, the unemployment rate fluctuated around 8% until early 1980s. The second breaks coincides with the emergence of a depressive cycle caused by the second oil-shock crisis. As it is shown, the hybrid two-stage procedure, that addresses uncertainty over the existence of the break and uncertainty over the autocorrelation properties of the series, is very effective in identifying important structural changes in the series so that the researcher can make valid economic inferences.

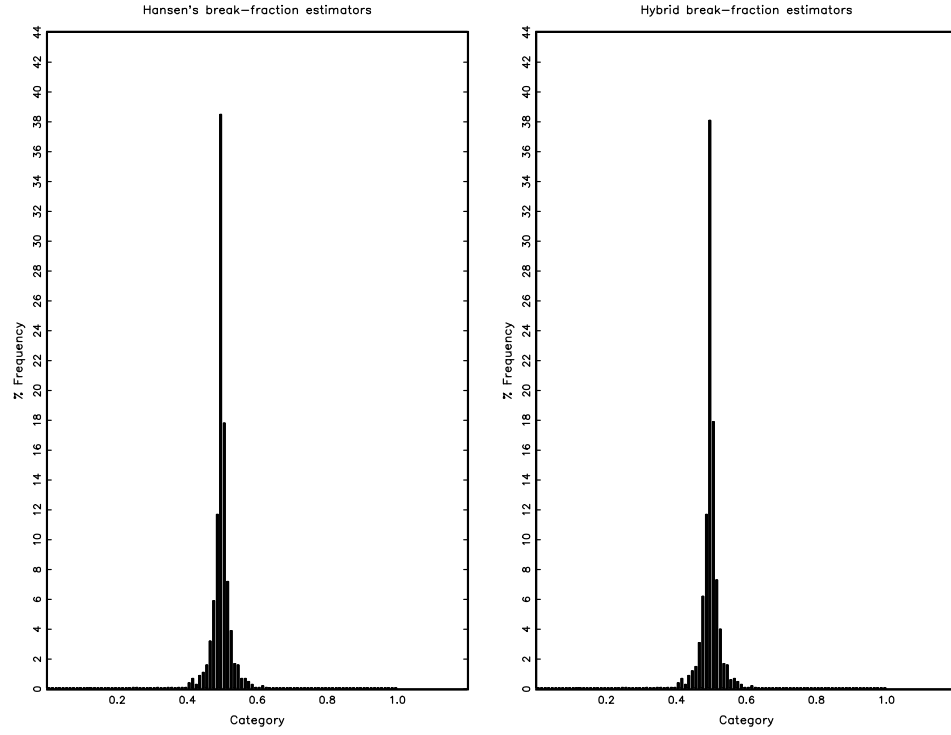
### 3.6 Conclusion

In summary, we have examined the finite sample performance of break size estimators that deal with both the uncertainty over the autoregressive properties of the data and uncertainty over the existence of a break. Since knowledge about the exact autoregressive properties of the stochastic component is typically limited, we design a two-stage procedure that works well without having to defend a potentially wrong view on the data's order of integration. At the same time, we employ weighted average estimators to account for the fact that breaks are not always obvious and often uncertainty whether a change has taken place in the behaviour of the series exists. In particular, we extended previous work by Hansen (2009), that used weighted average estimators, by using quasi-differenced or first differenced data depending on the selected autoregressive parameter that minimises the sum of squared residuals across a number of options, following the approach developed by Harvey and Leybourne (2013). It was found that the approach works well and achieves accurate predictions about the break size. The advantage of the procedure lies on the fact that it accurately predicts the break fraction, if it exists, so that it subsequently estimates the break size with great precision, irrespective of the autocorrelation properties of the series. We showed that the hybrid estimators proposed in this article goes a long way in improving Hansen's unfiltered approach and

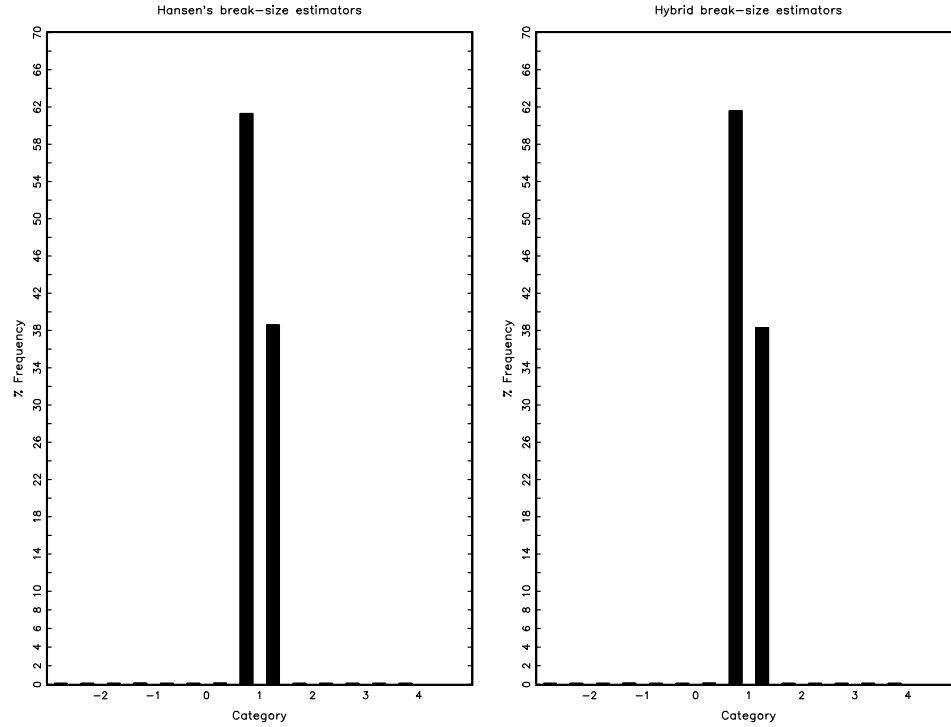
should therefore have practical appeal. The empirical example provided, picks up all the important changes the U.S. unemployment series has gone through over the period examined.



### 3.7 Figures

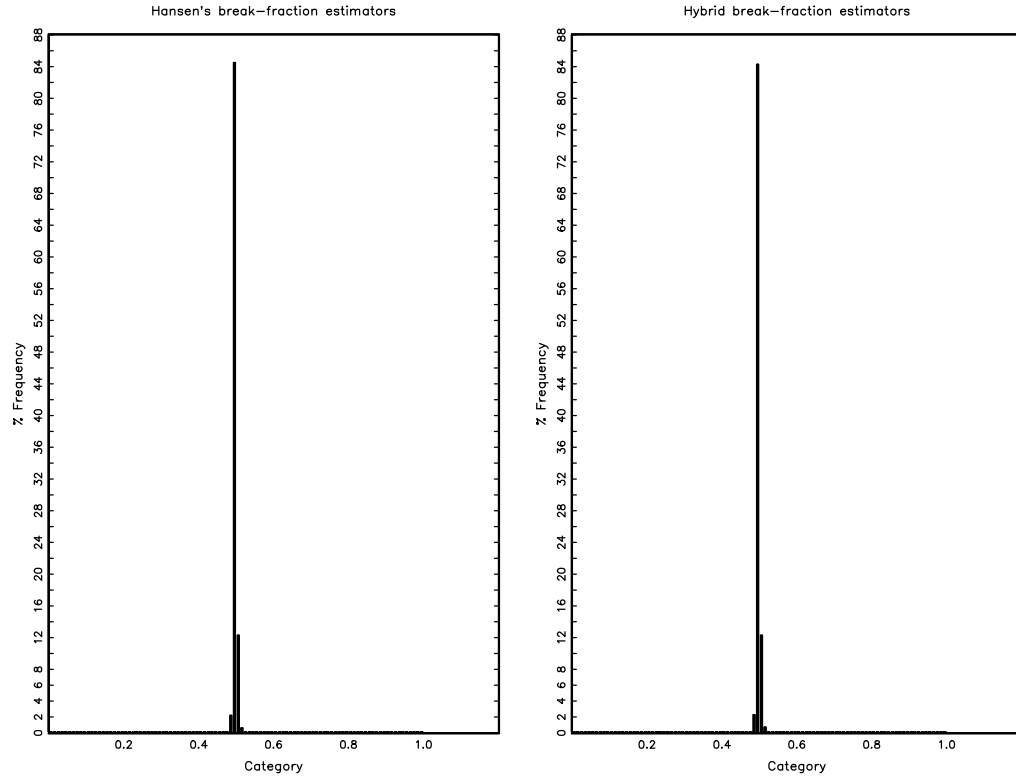


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

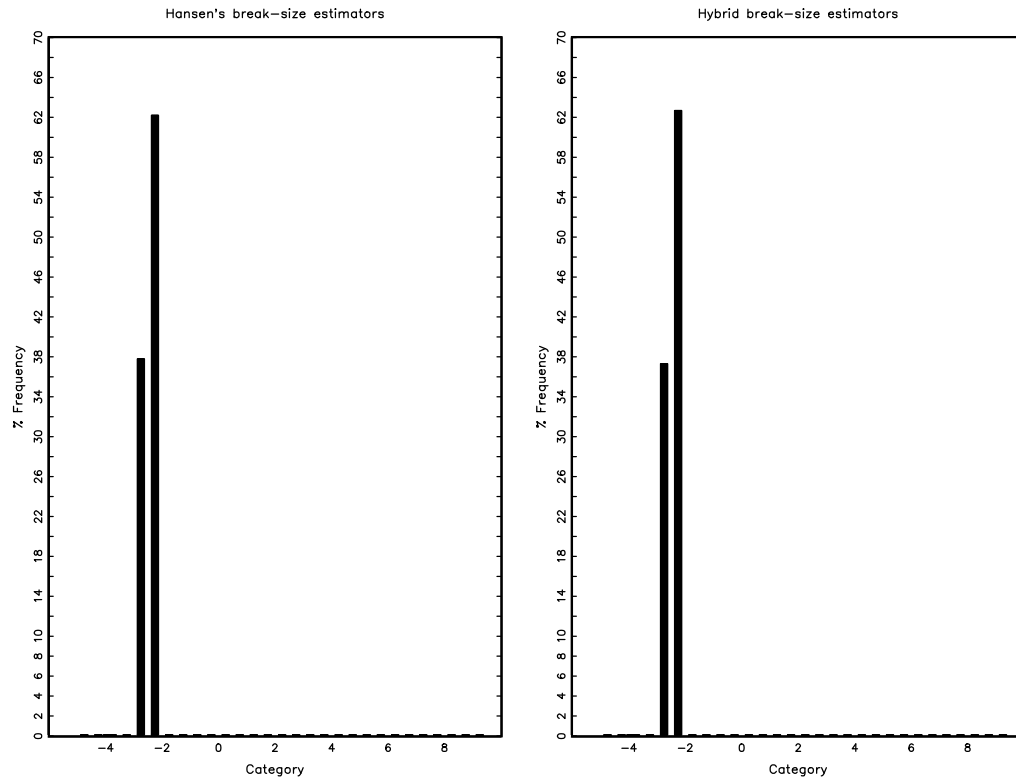


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-1: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0$ ,  $\gamma_1 = 1$ .

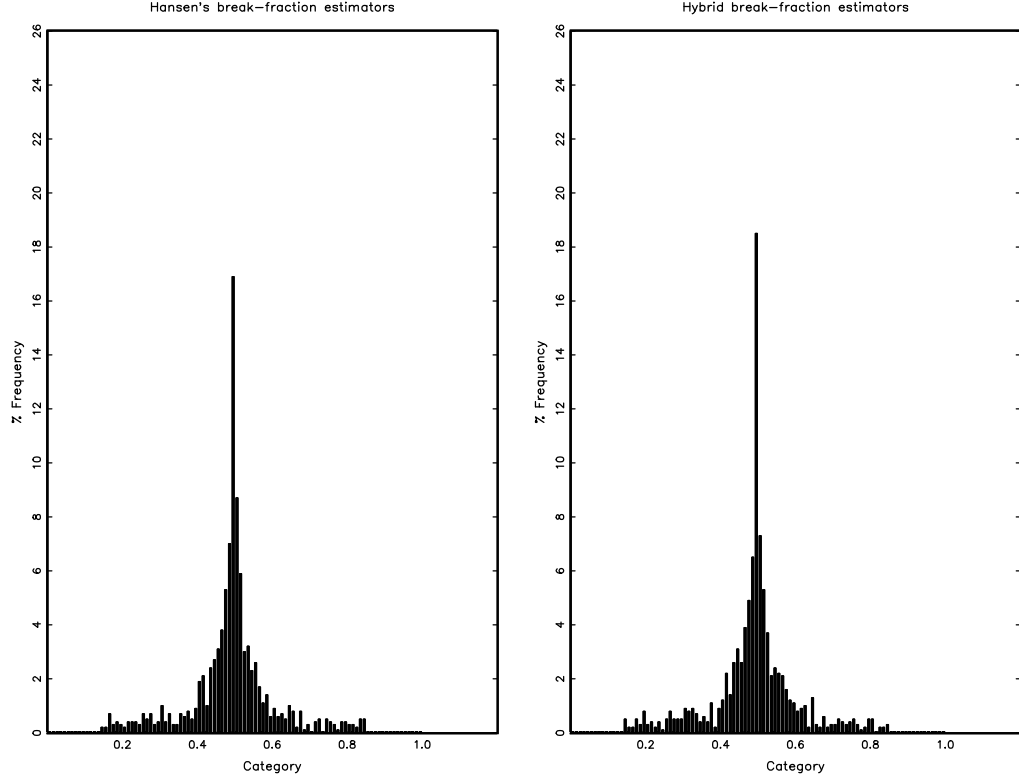


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

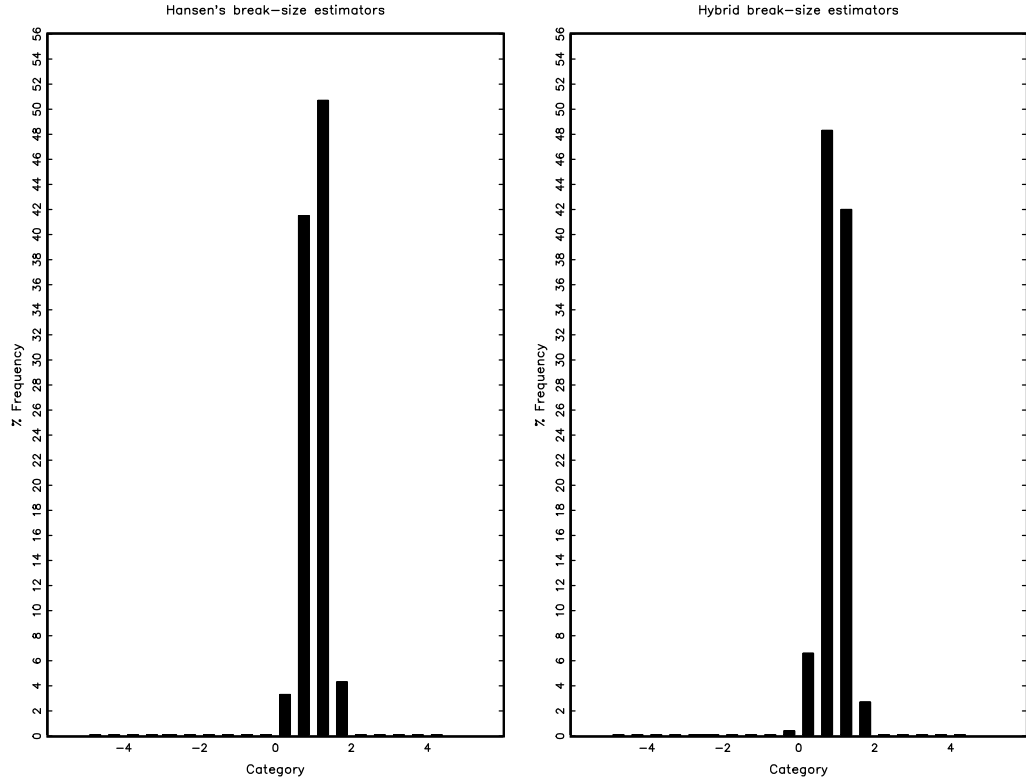


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-2: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0$ ,  $\gamma_1 = -2.5$ .

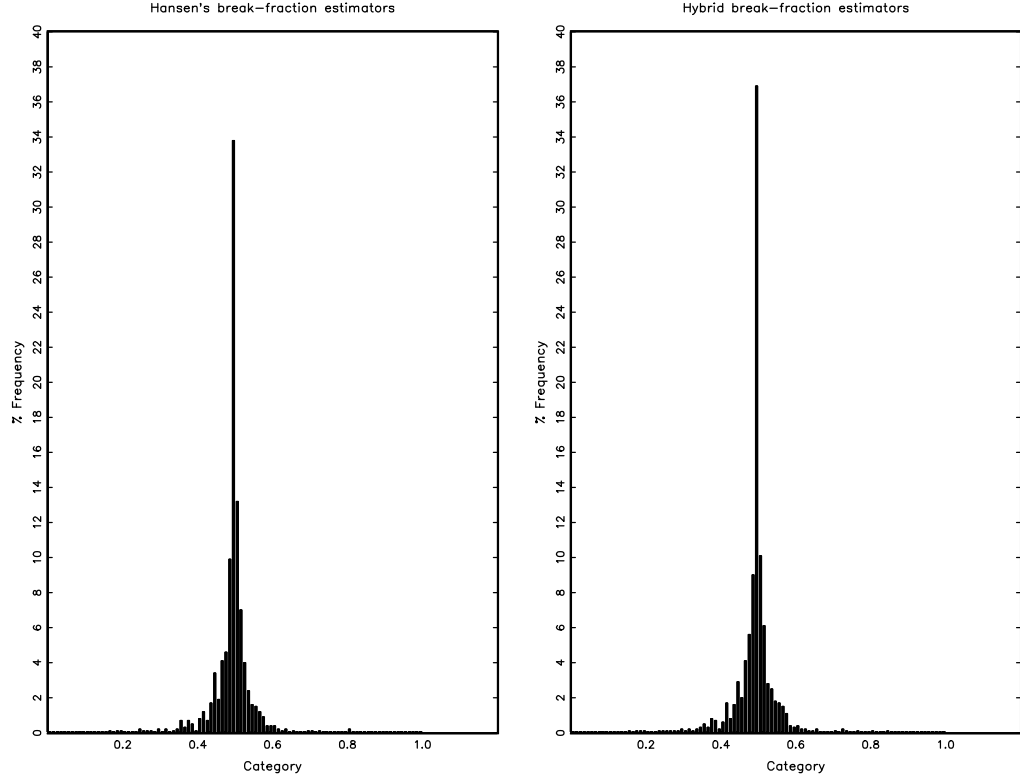


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

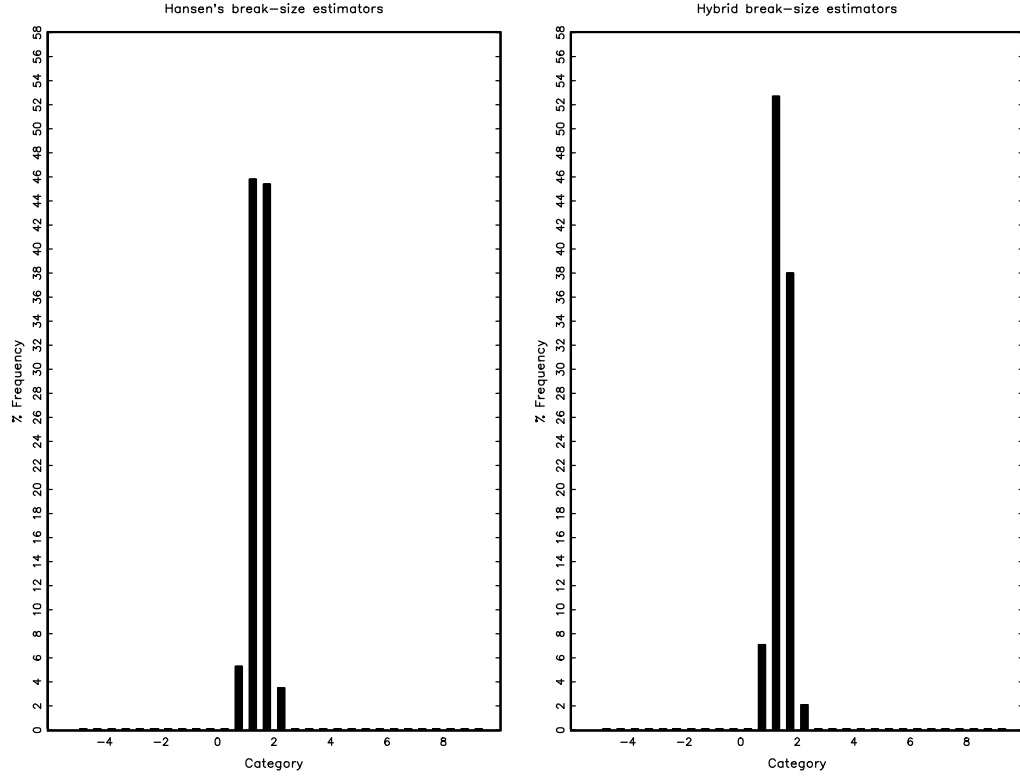


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-3: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0.5$ ,  $\gamma_1 = 1$ .

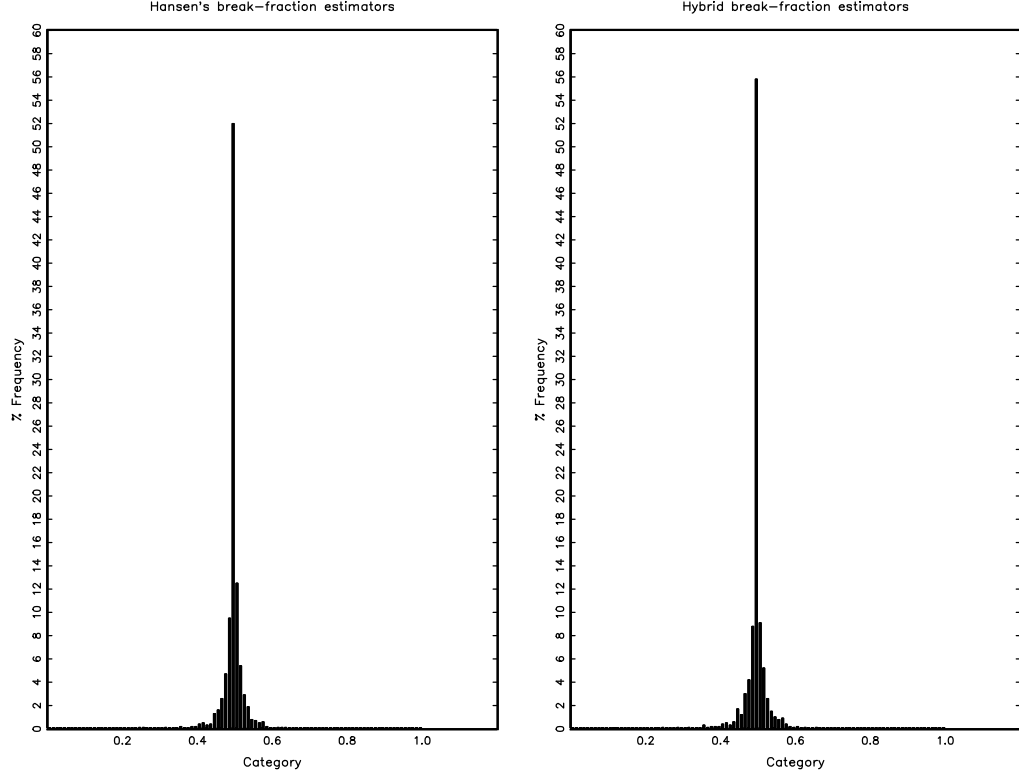


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

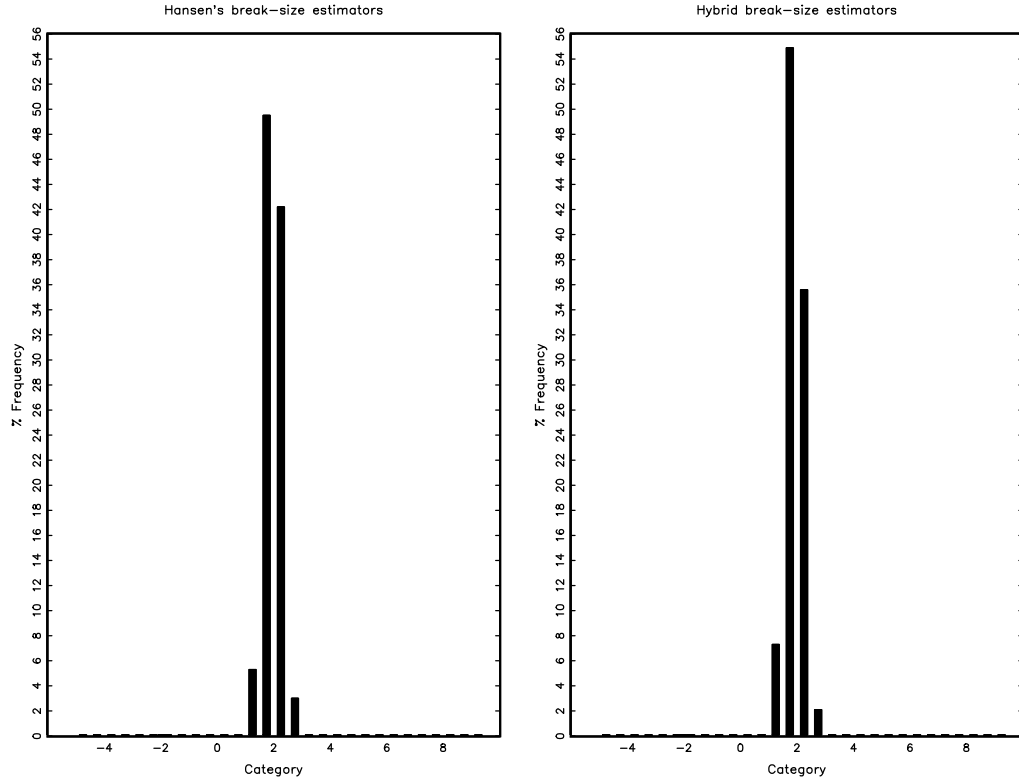


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-4: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0.5$ ,  $\gamma_1 = 1.5$ .

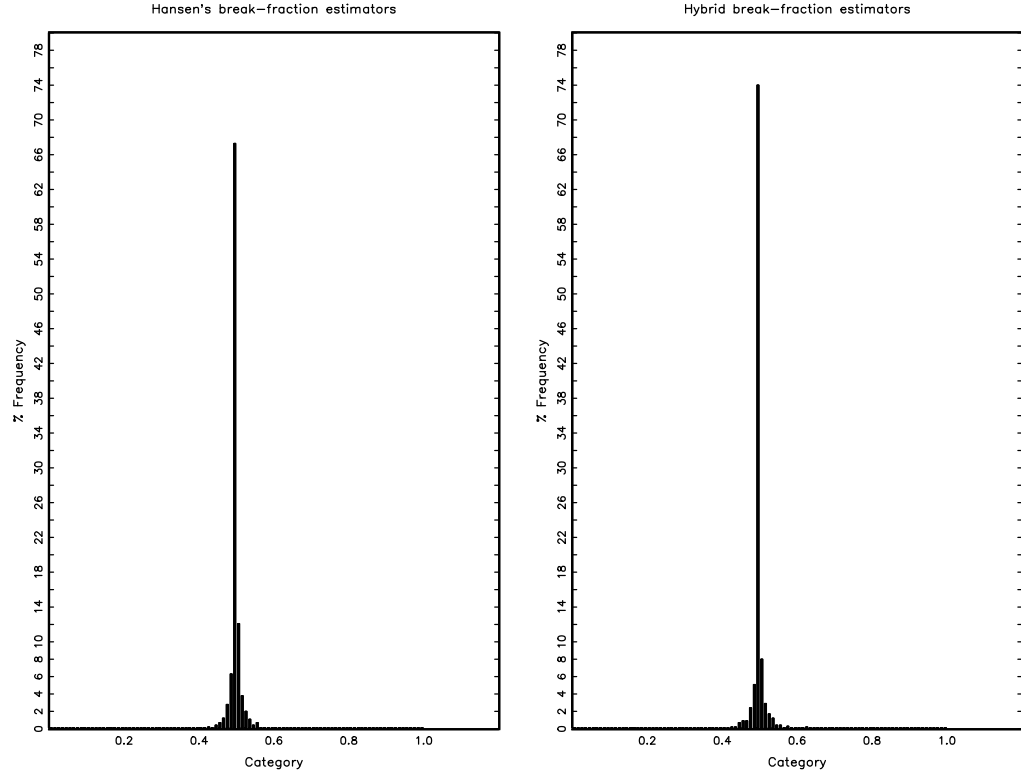


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

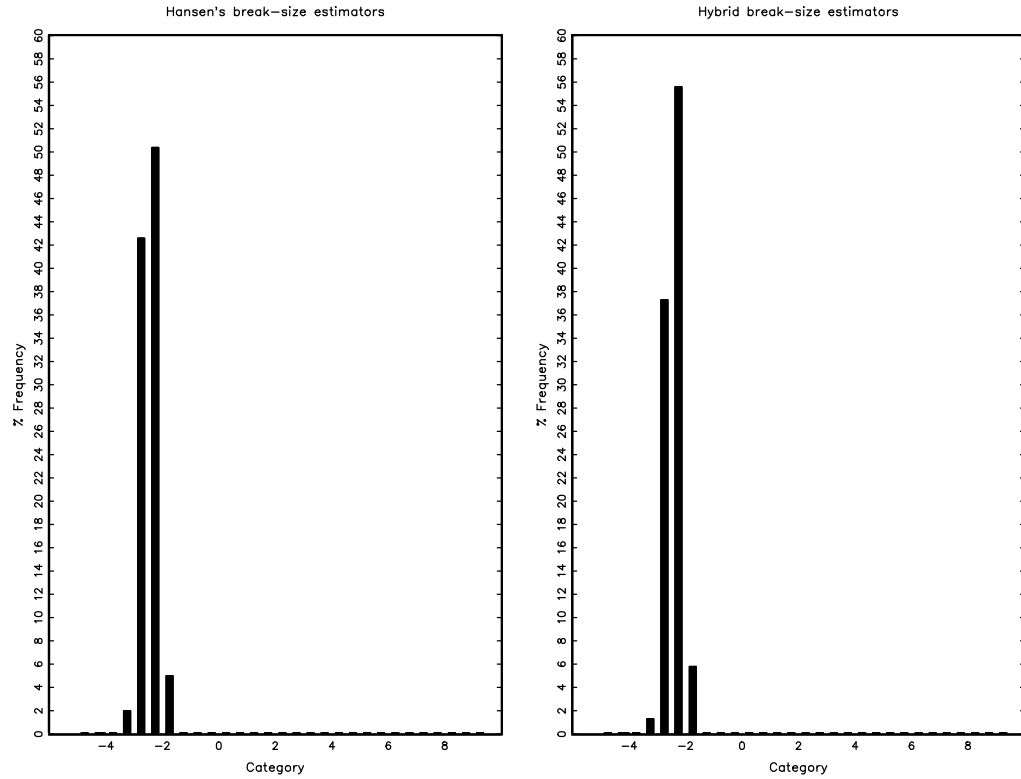


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-5: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0.5$ ,  $\gamma_1 = 2$ .

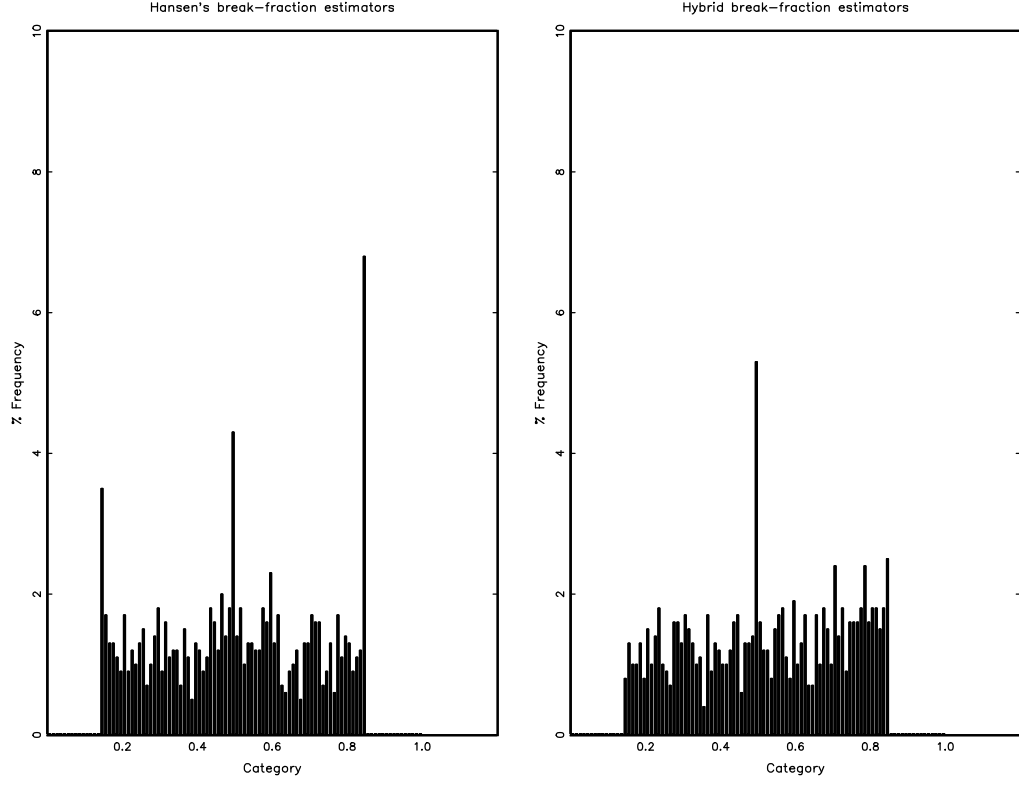


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

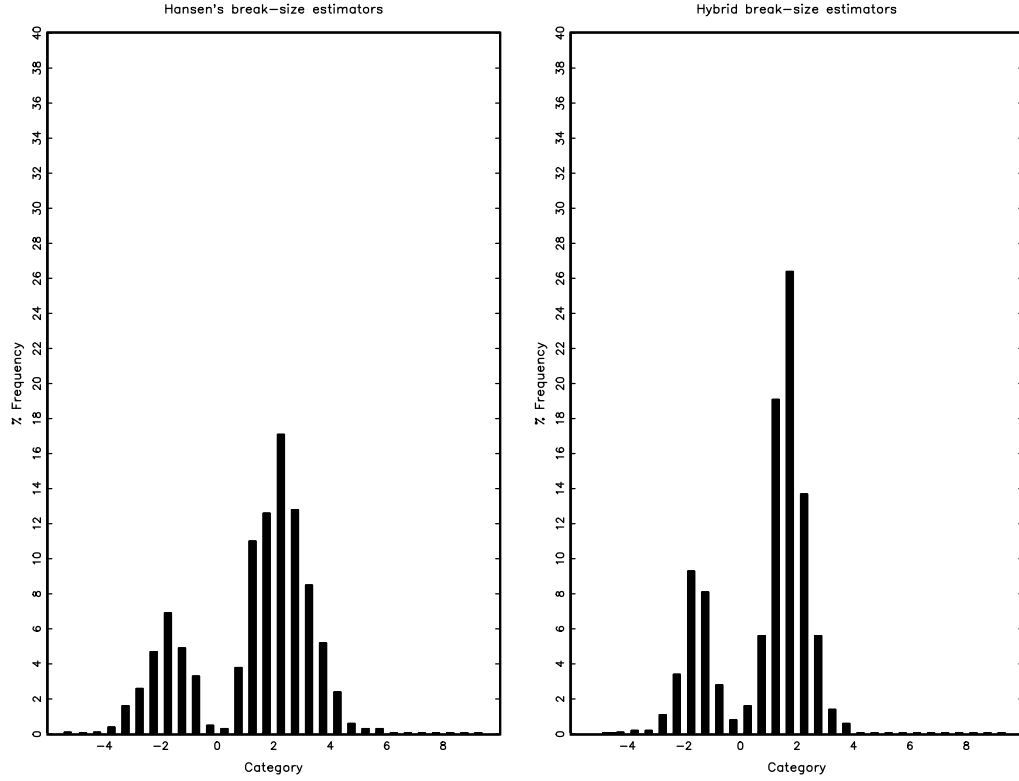


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-6: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0.5$ ,  $\gamma_1 = -2.5$ .

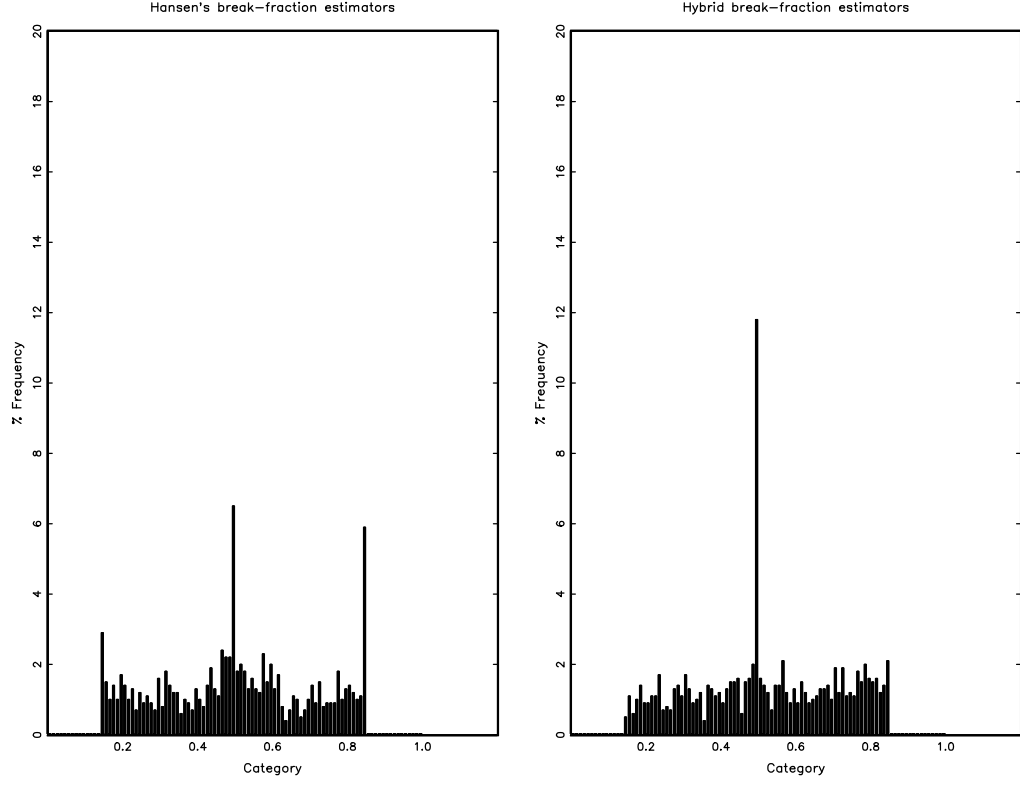


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

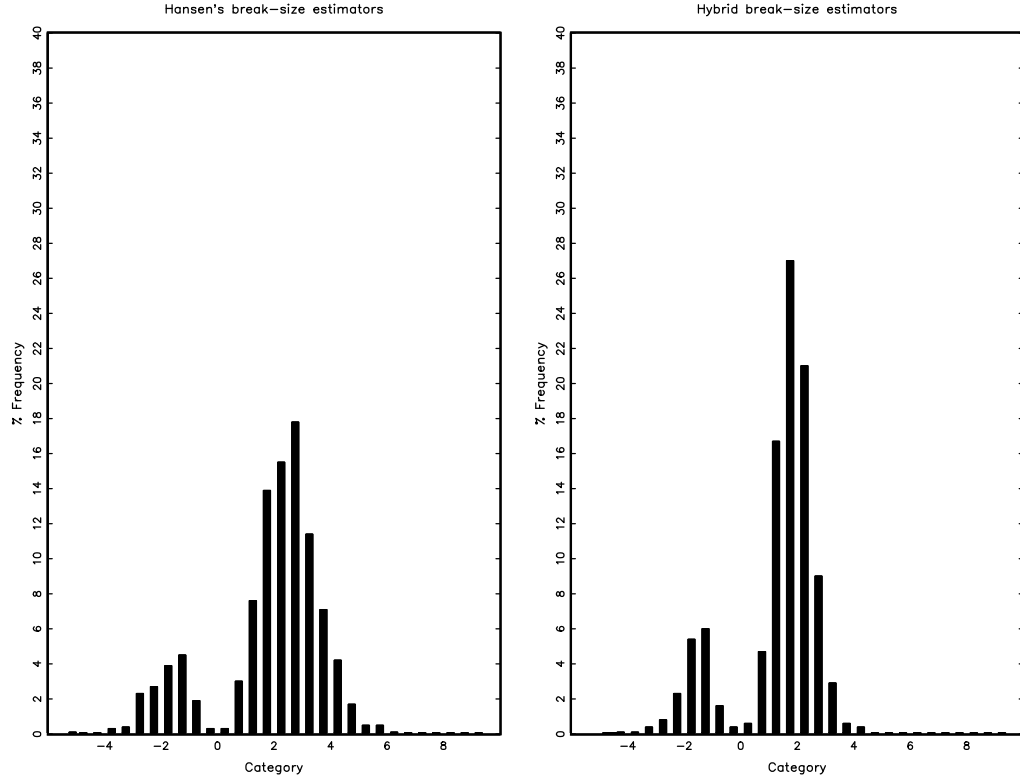


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-7: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0.9$ ,  $\gamma_1 = 1$ .



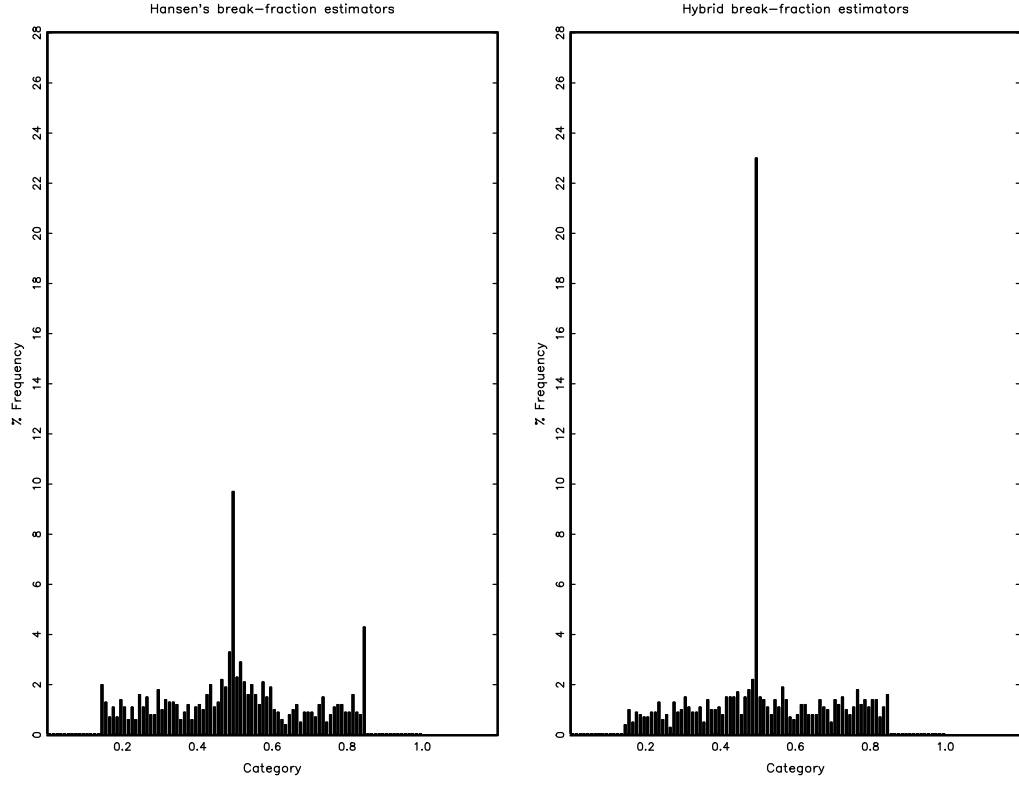
a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches



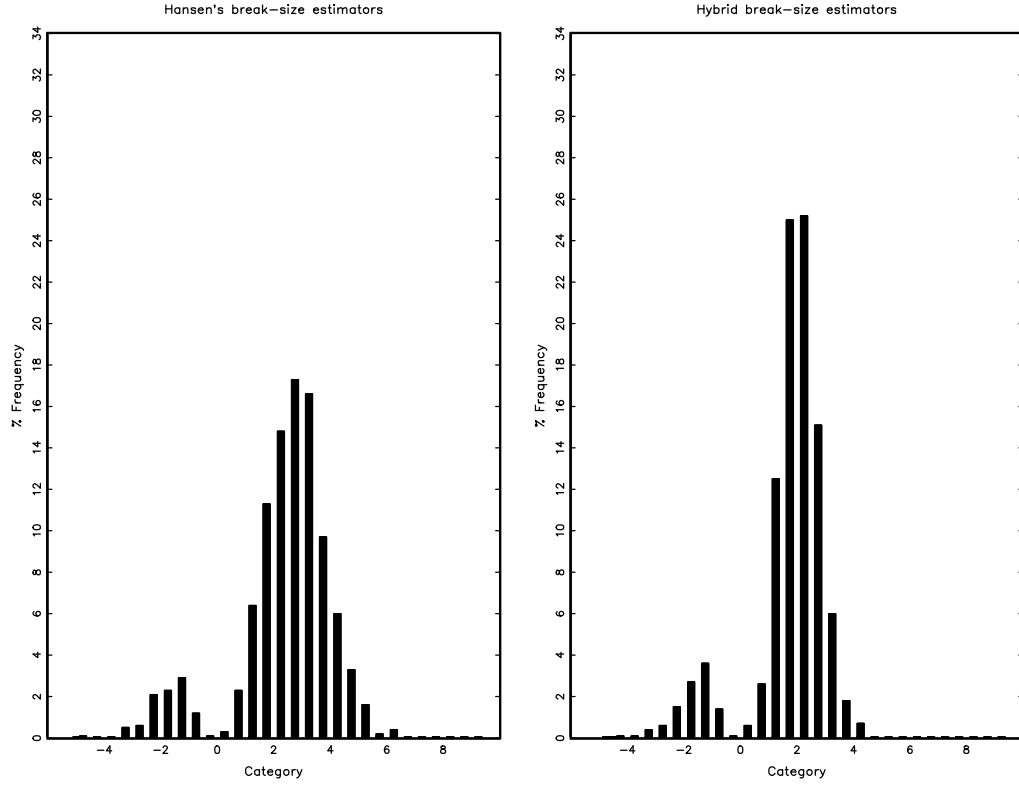
b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-8: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0.9$ ,  $\gamma_1 = 1.5$ .



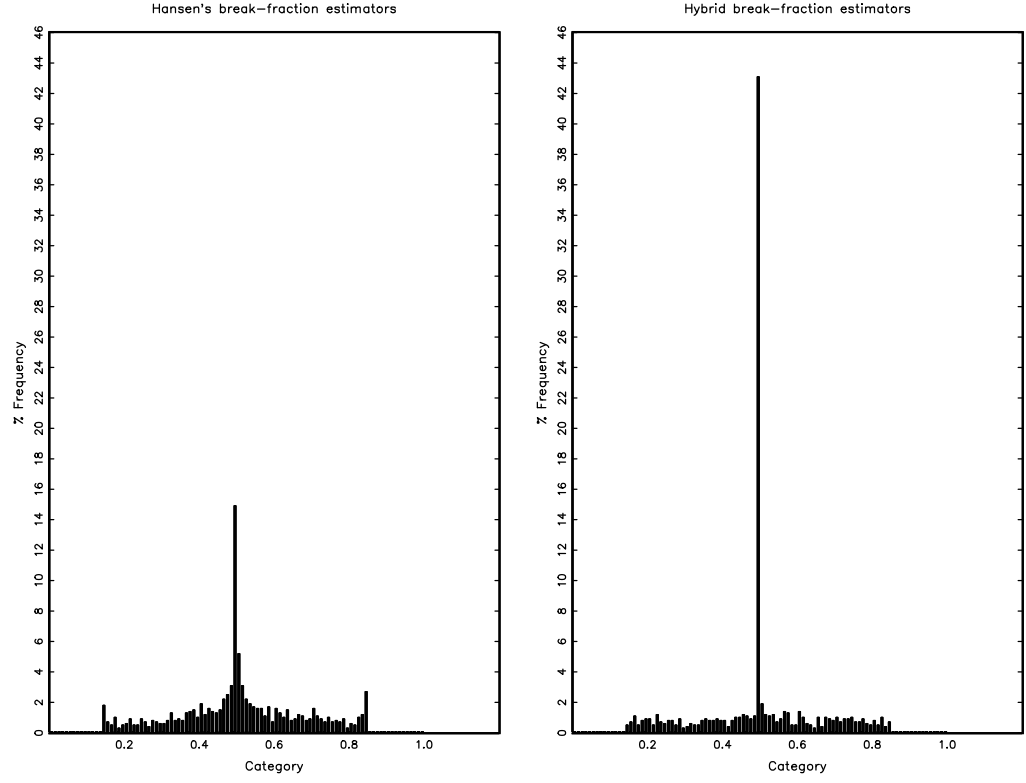


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

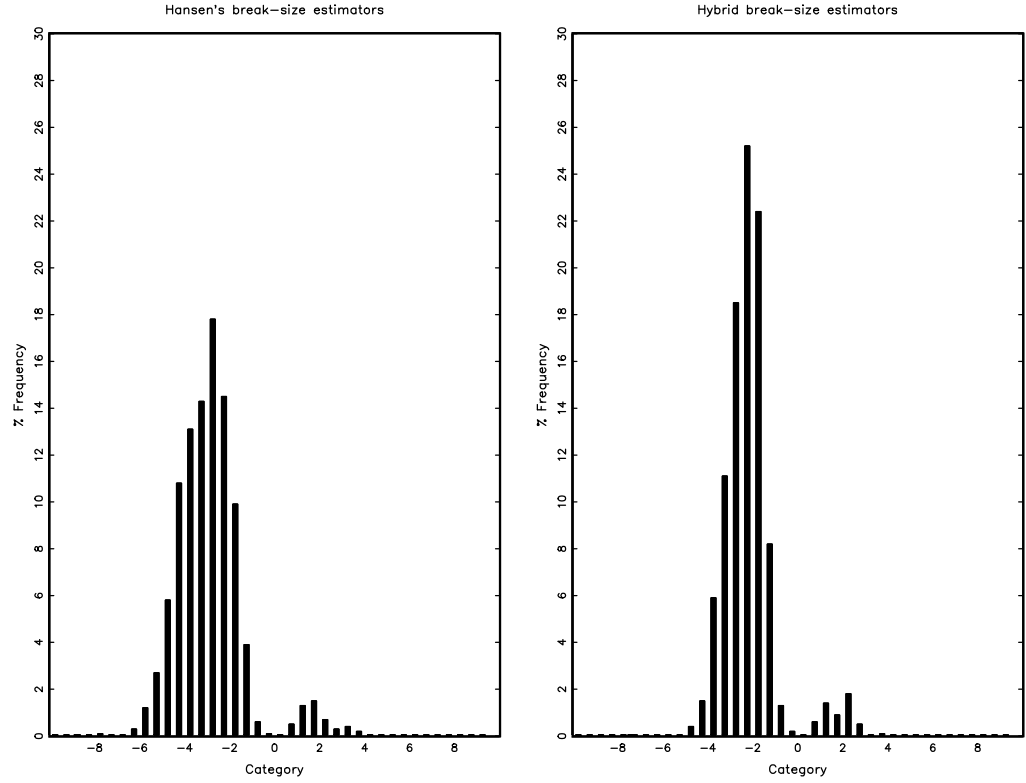


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-9: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0.9$ ,  $\gamma_1 = 2$ .

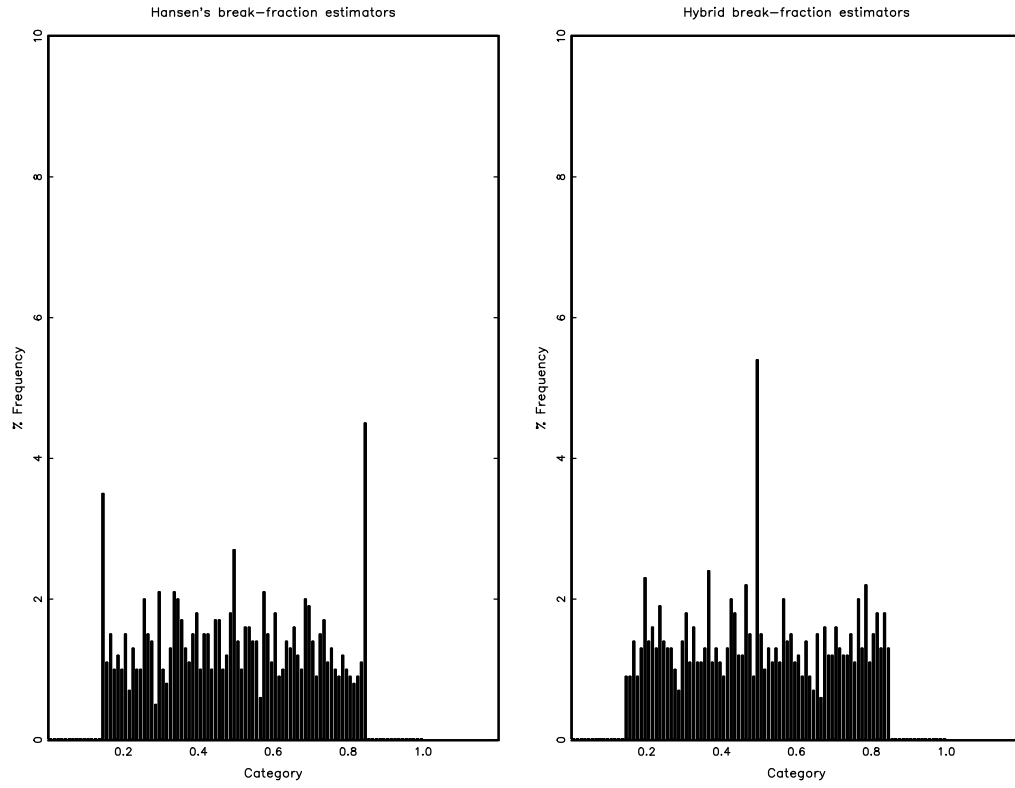


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

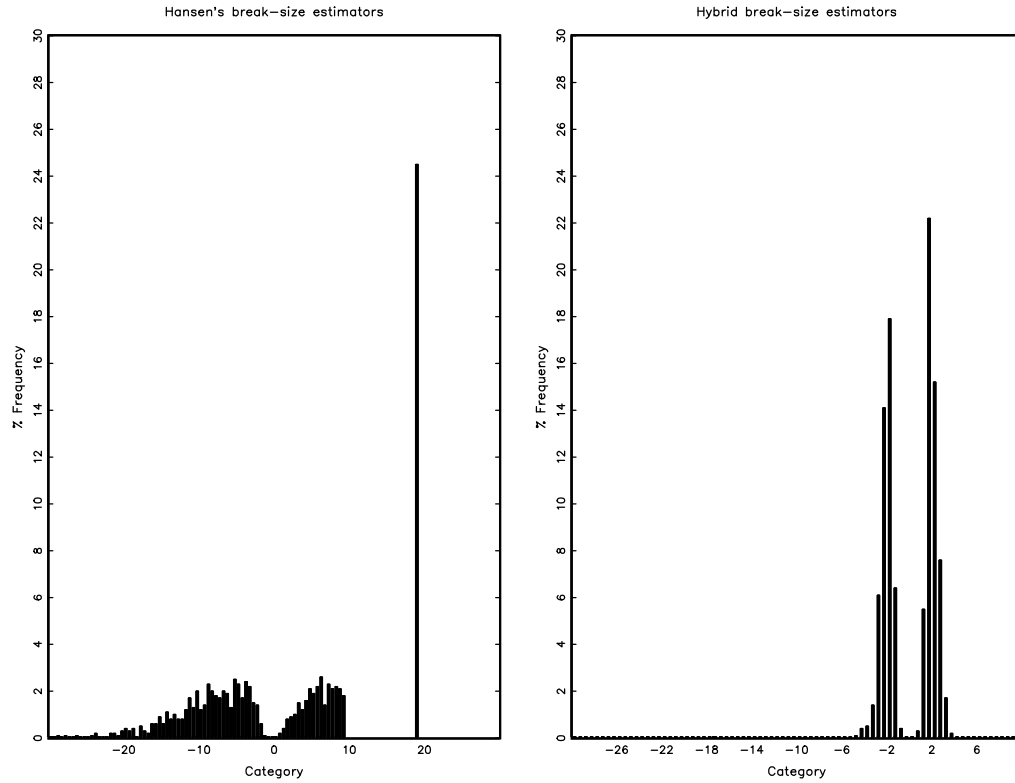


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-10: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 0.9$ ,  $\gamma_1 = -2.5$ .

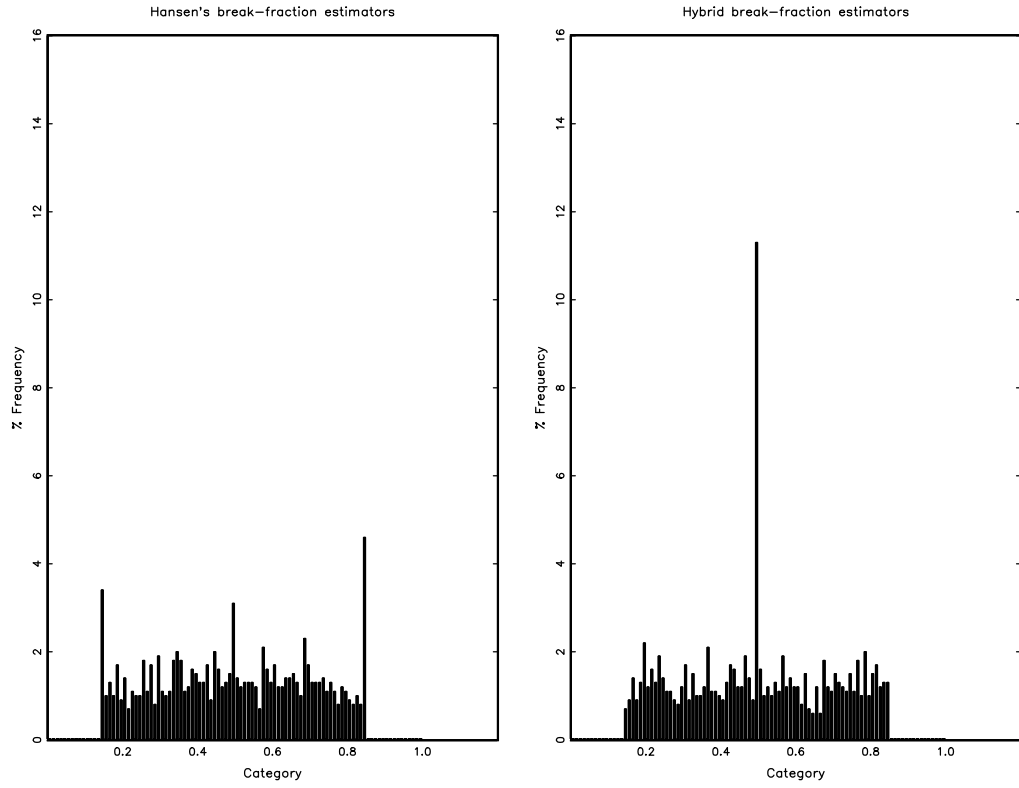


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

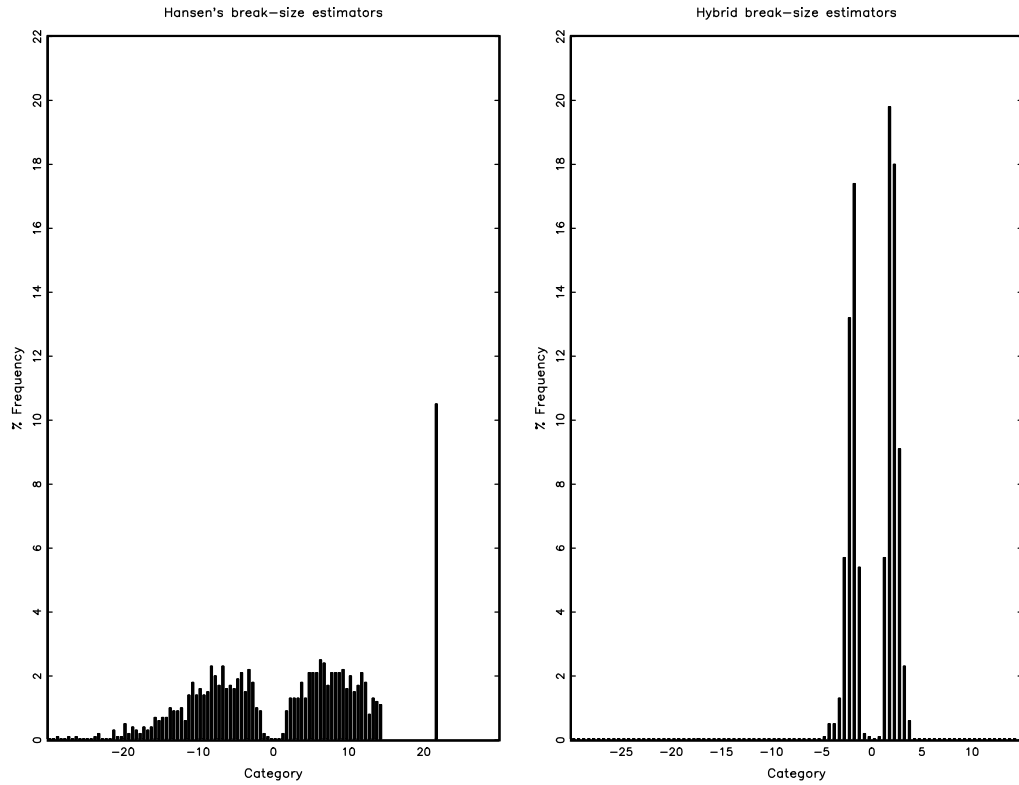


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-11: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 1$ ,  $\gamma_1 = 1$ .

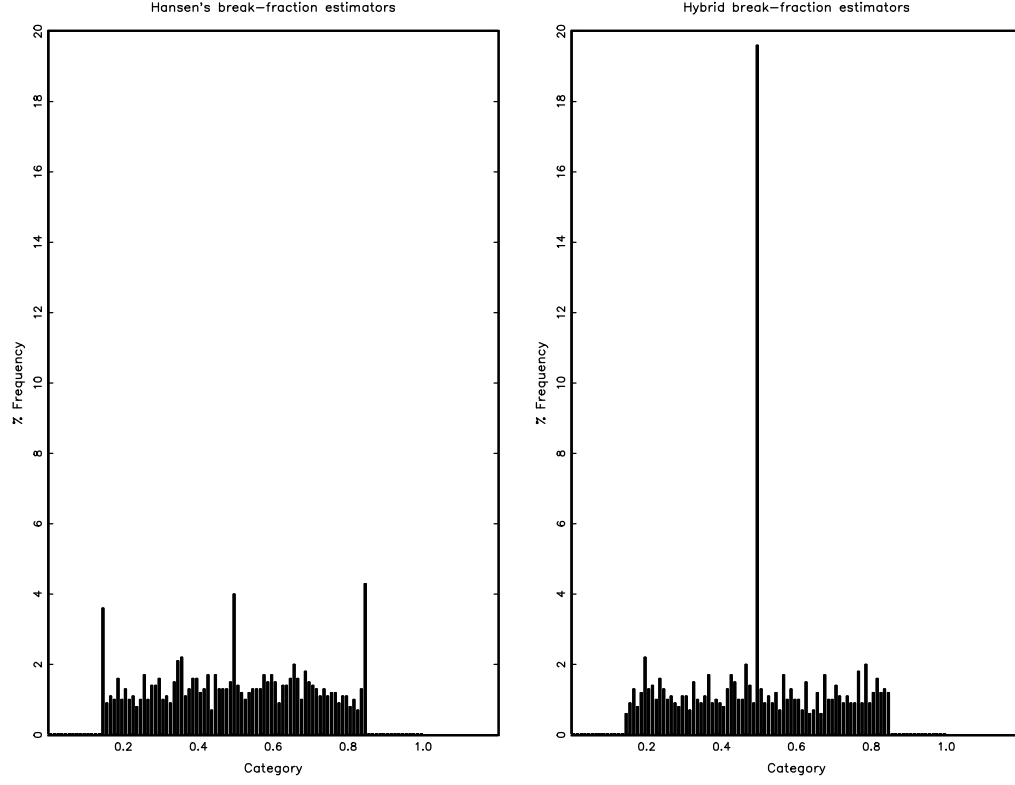


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

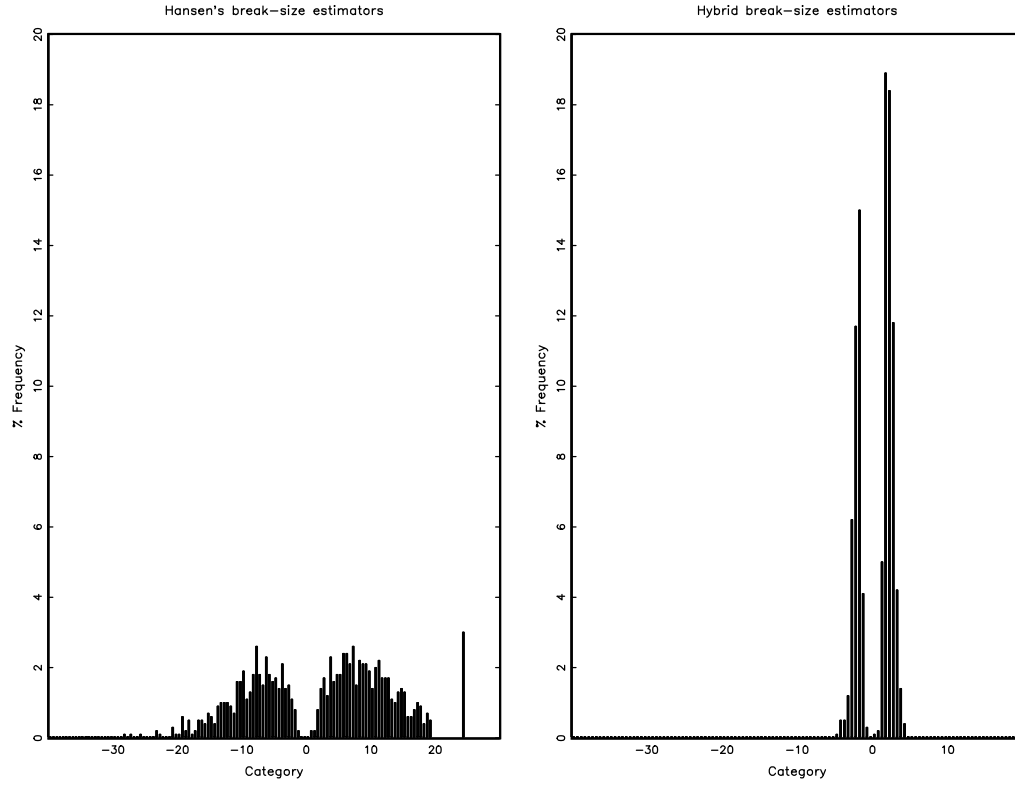


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-12: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 1$ ,  $\gamma_1 = 1.5$ .

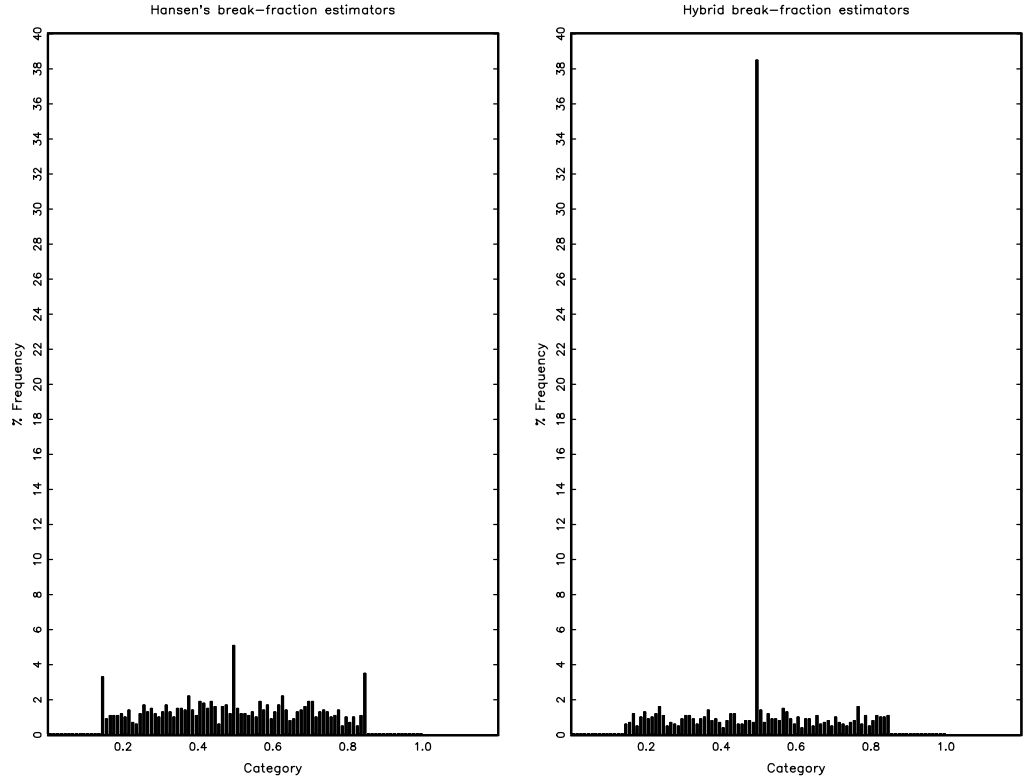


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

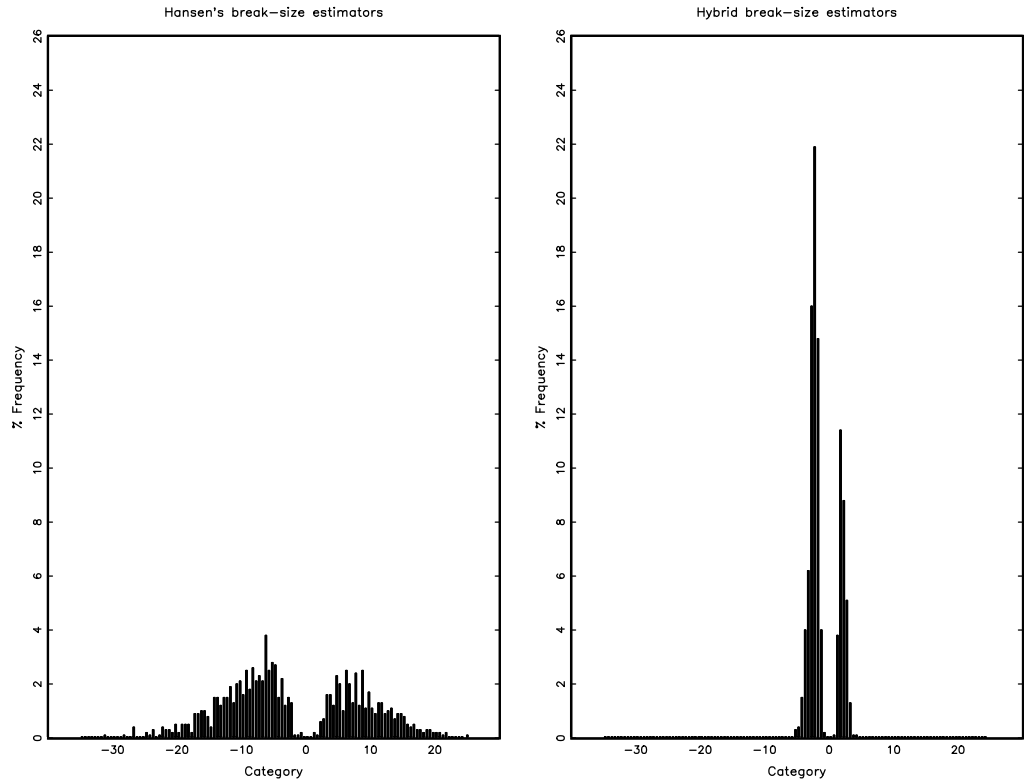


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-13: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 1$ ,  $\gamma_1 = 2$ .

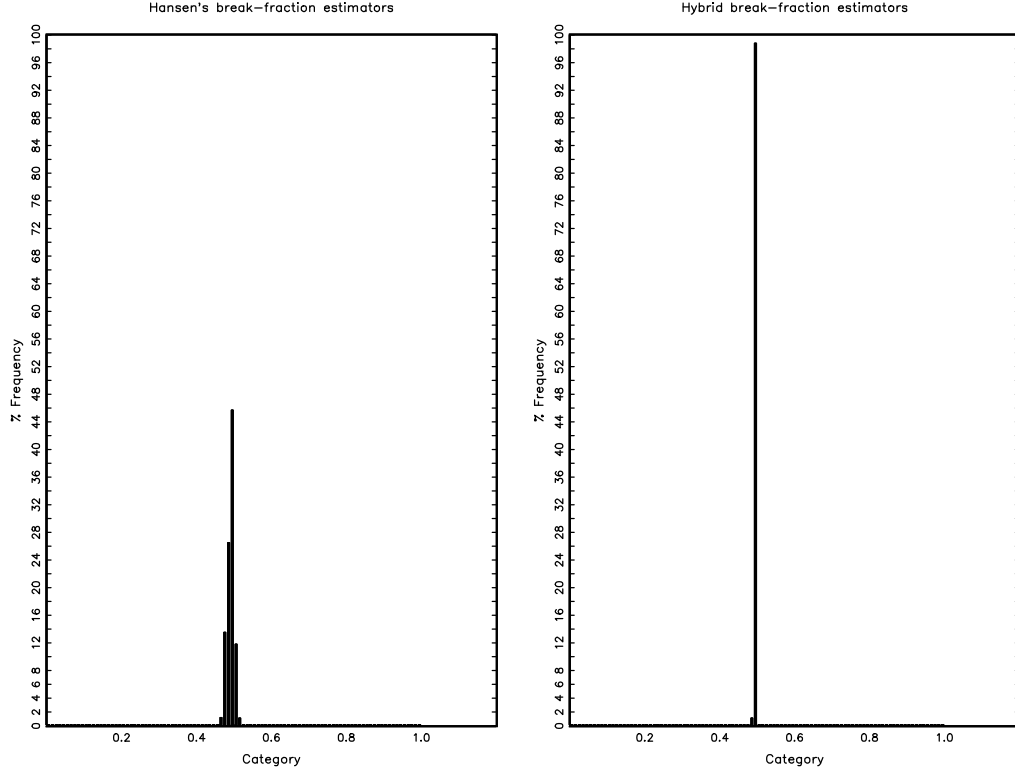


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

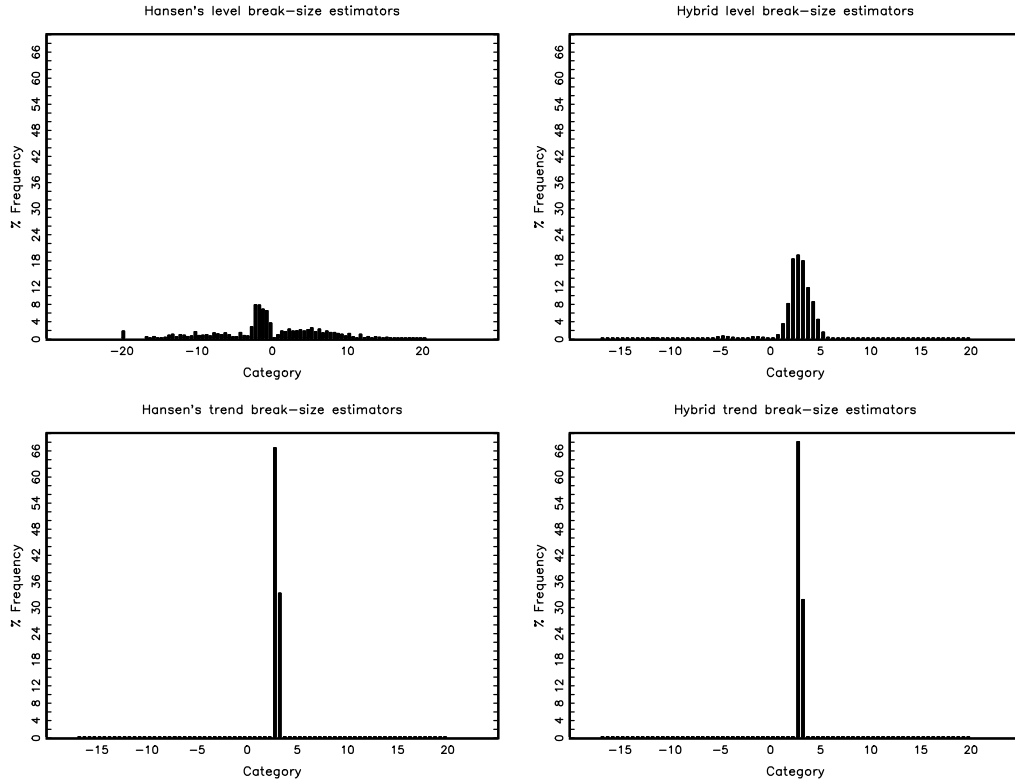


b) Histograms of the level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-14: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,  $\rho = 1$ ,  $\gamma_1 = -2.5$ .



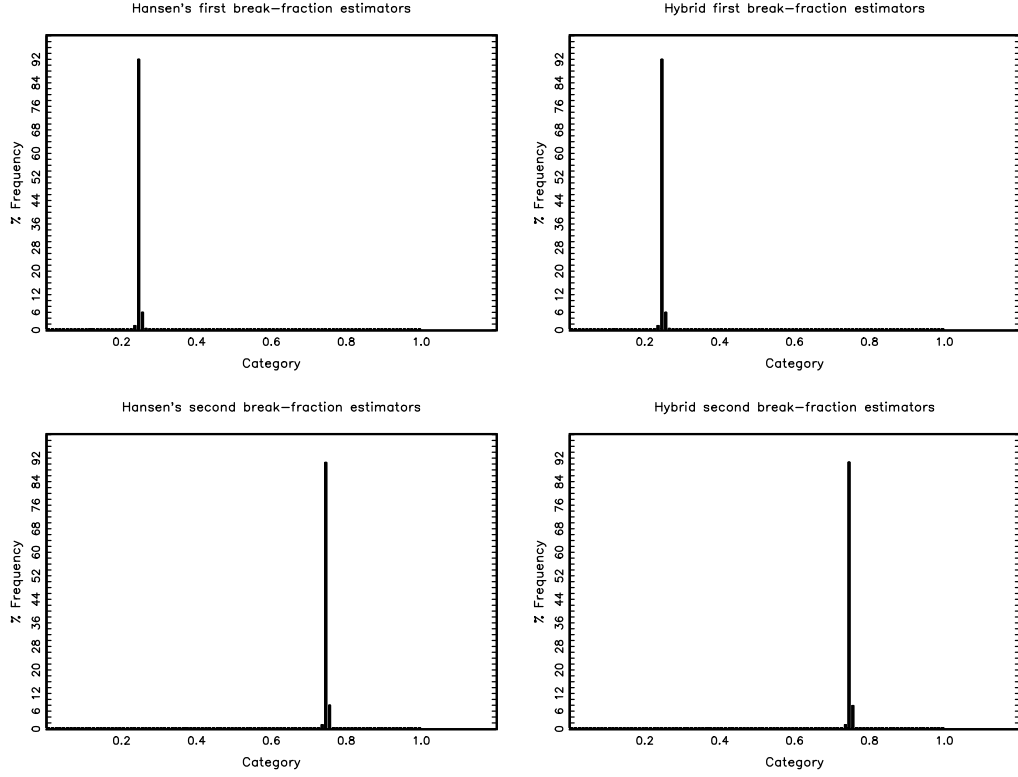
a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches



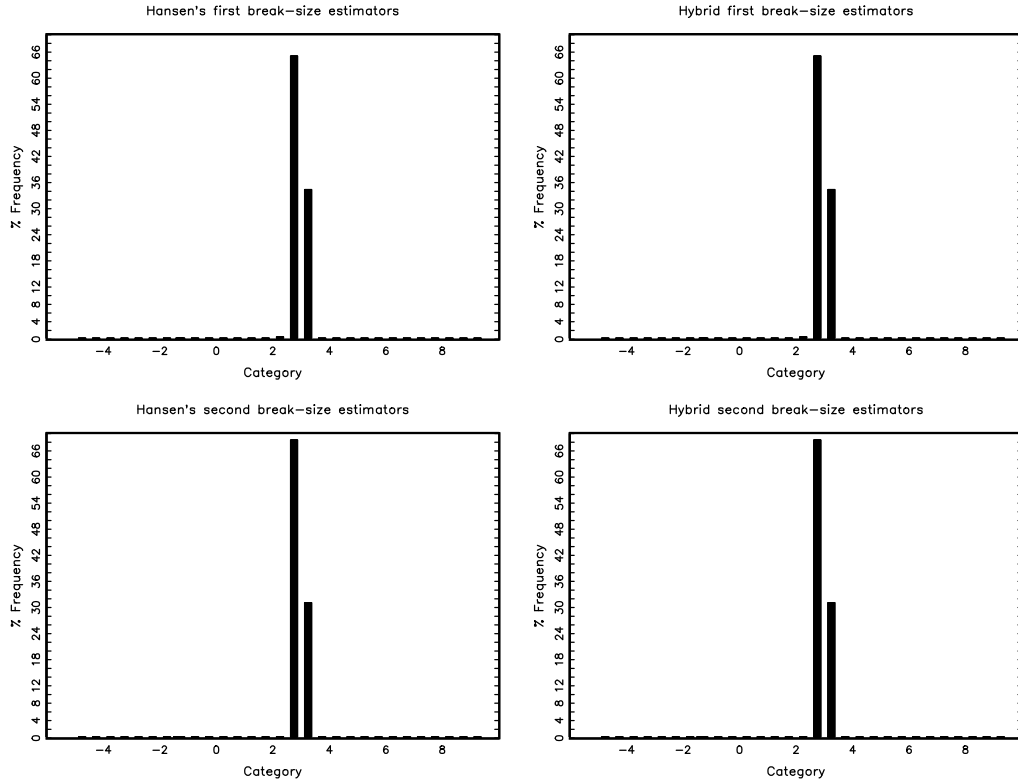
b) Histograms of the level and trend break-size estimators as given by Hansen and Hybrid approaches

Figure 3-15: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau^* = 0.5$ ,

$$\rho = 1, \gamma_1 = 3, \gamma_2 = 3.$$



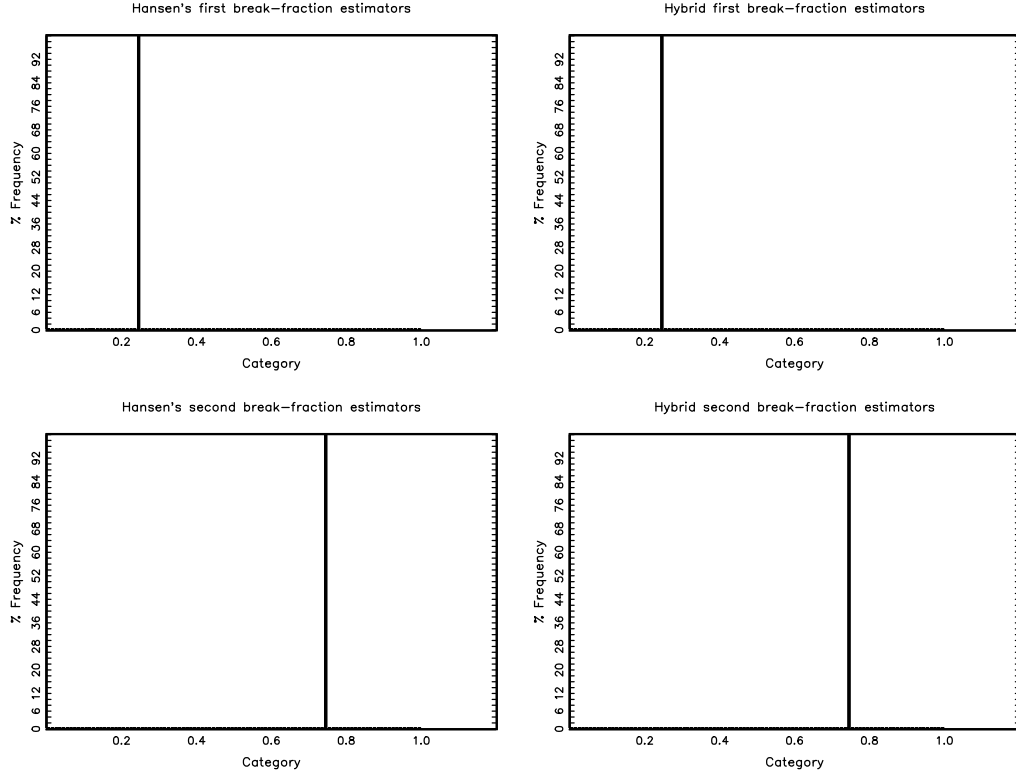
a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches



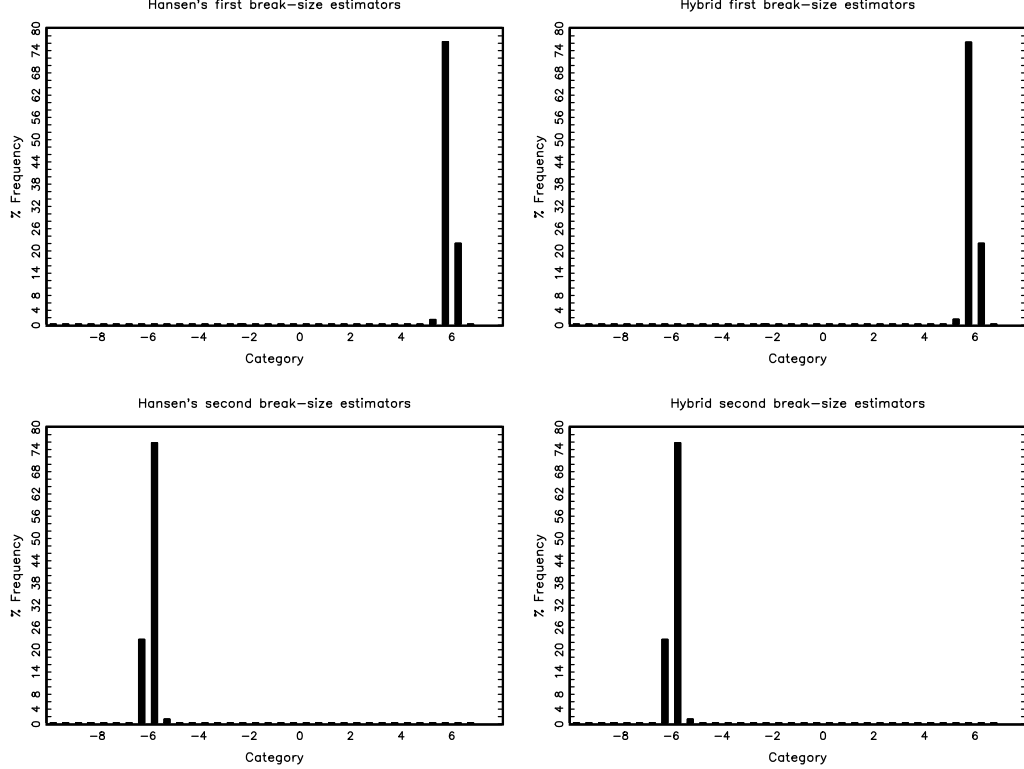
b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-16: Finite sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 0$ ,  $\gamma_1 = 3$ ,  $\gamma_2 = 3$ .





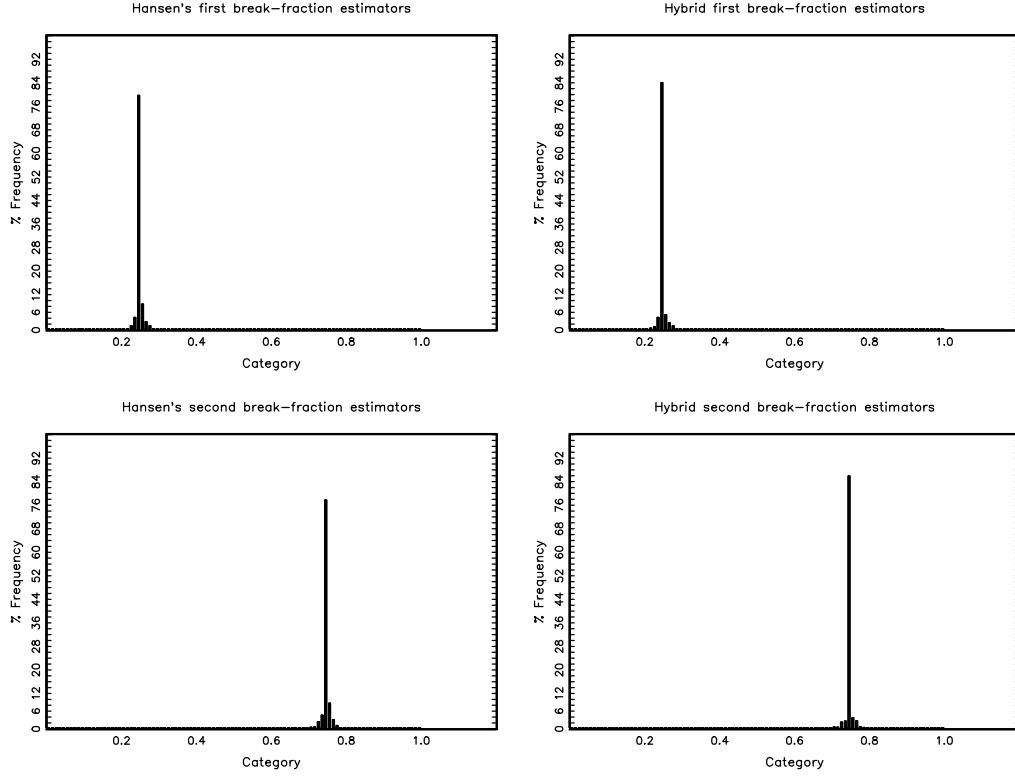
a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches



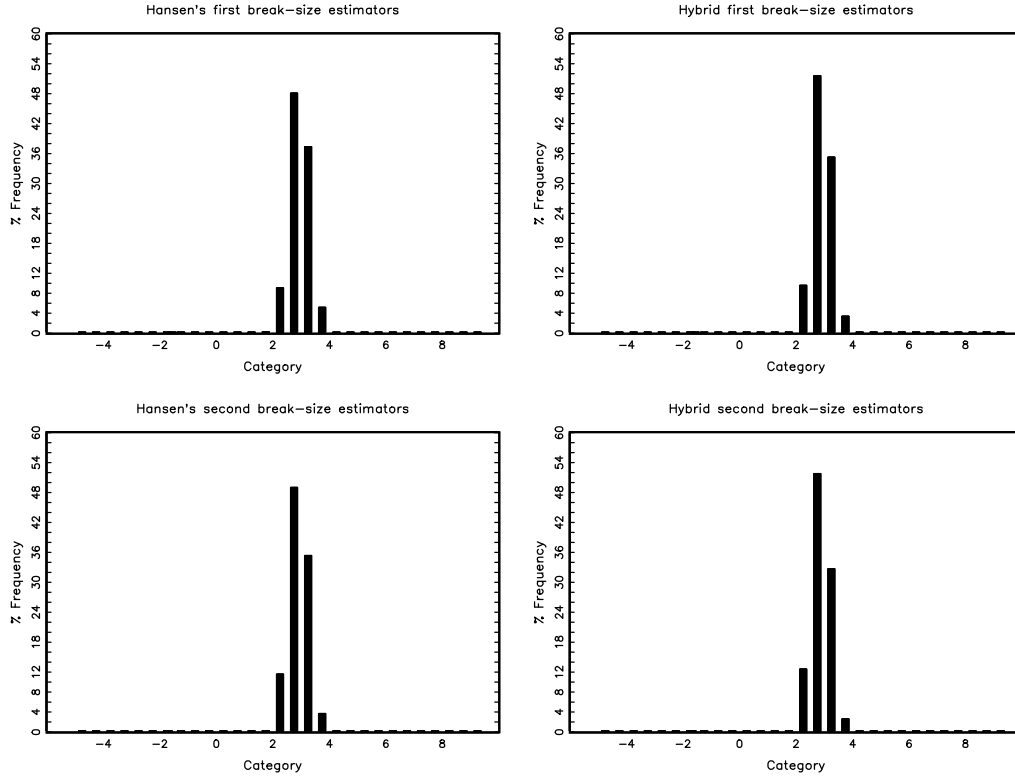
b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-17: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,

$\rho = 0$ ,  $\gamma_1 = 6$ ,  $\gamma_2 = -6$ .

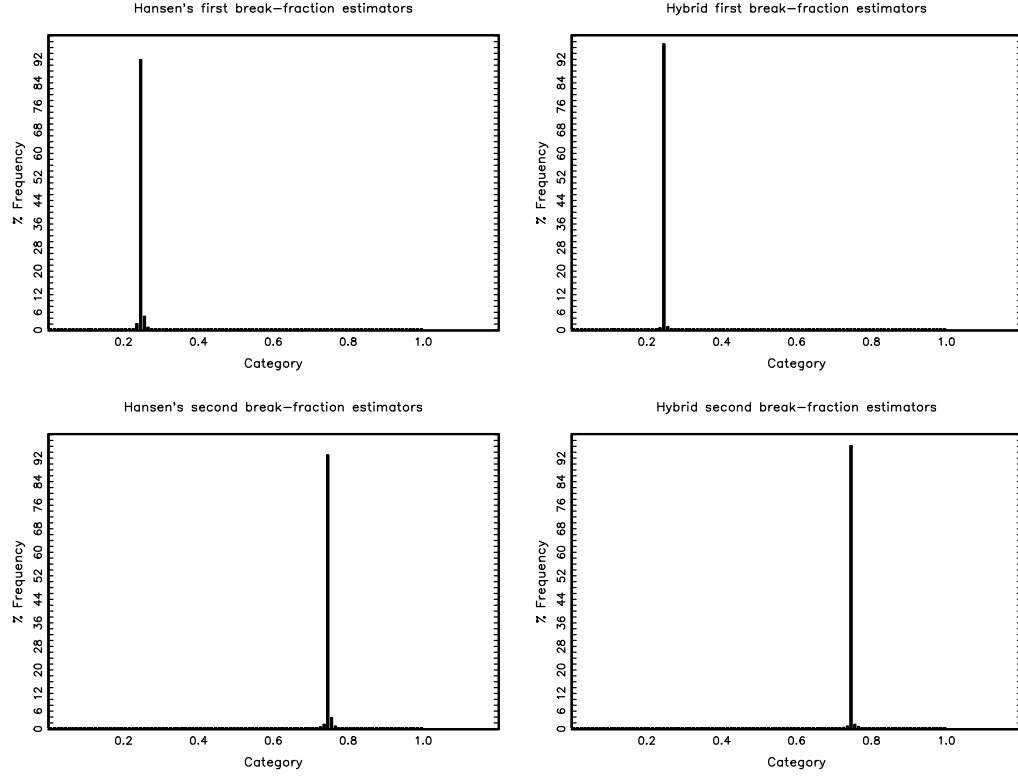


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

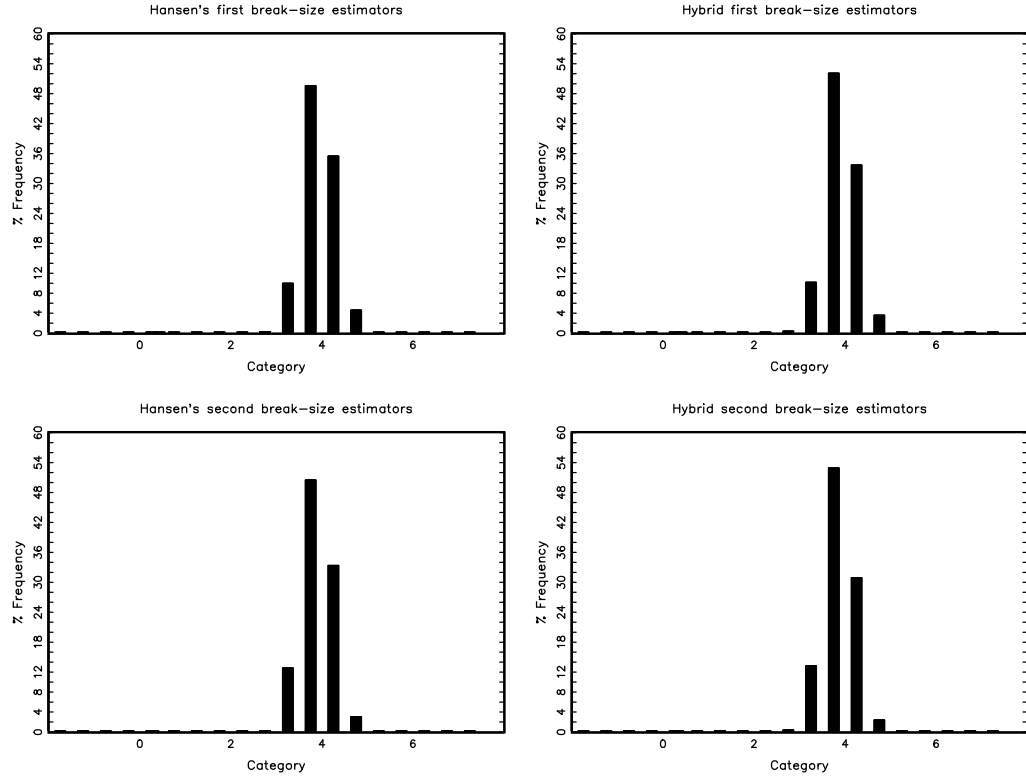


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-18: Finite sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 0.5$ ,  $\gamma_1 = 3$ ,  $\gamma_2 = 3$ .

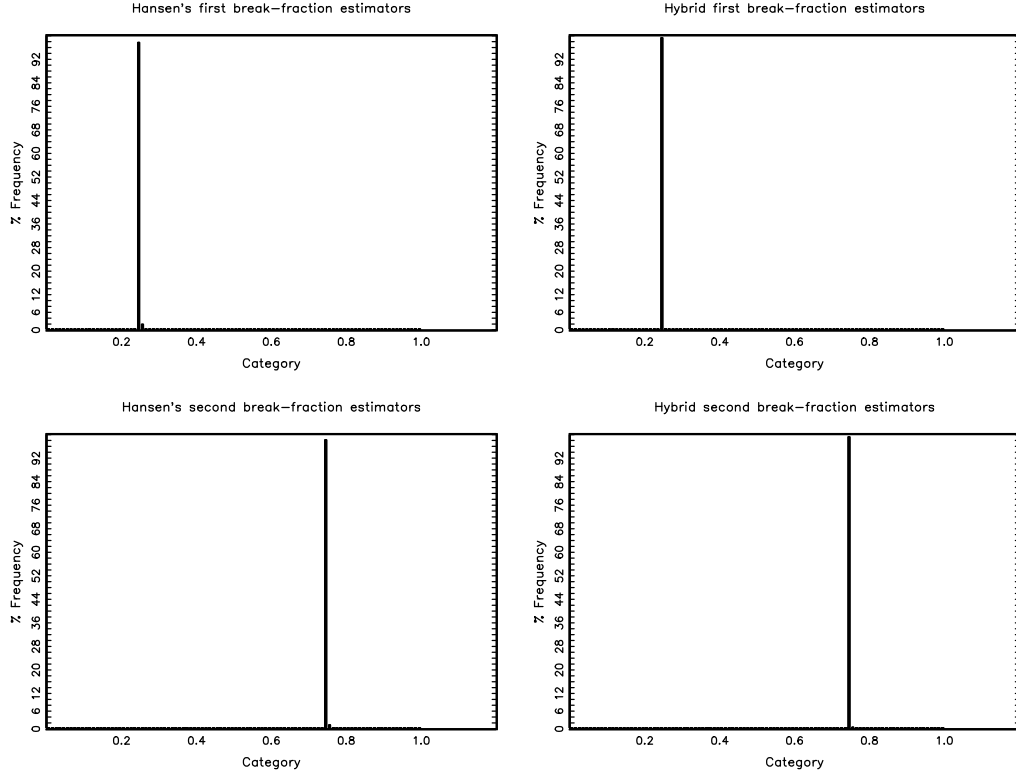


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

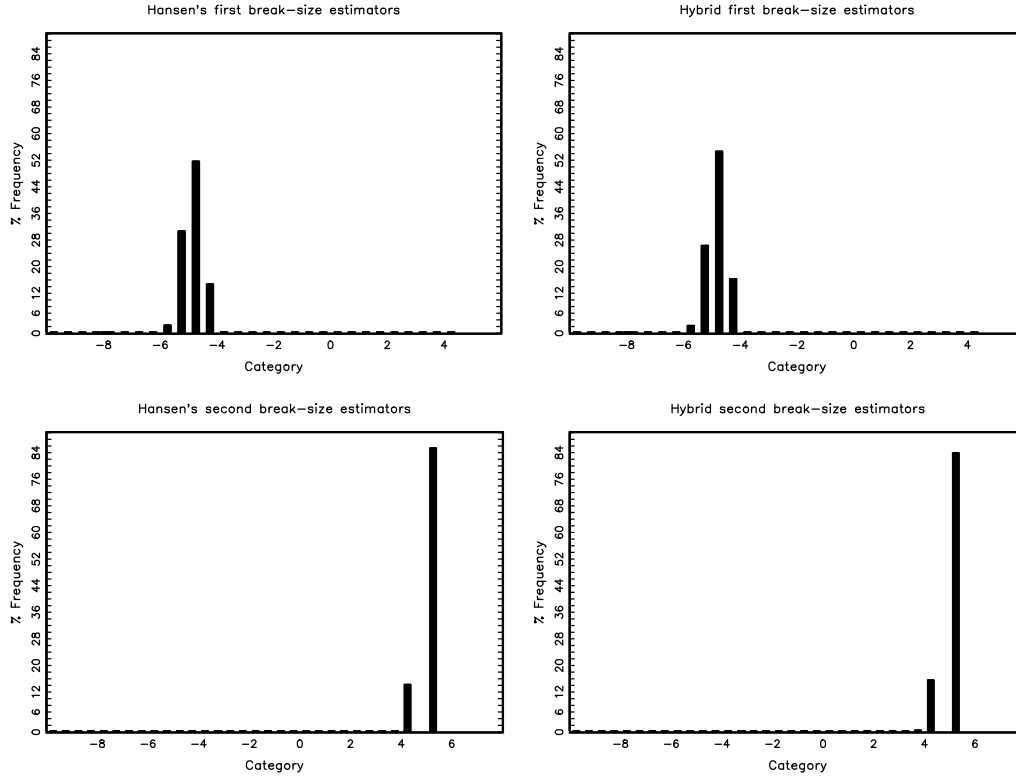


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-19: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 0.5$ ,  $\gamma_1 = 4$ ,  $\gamma_2 = 4$ .

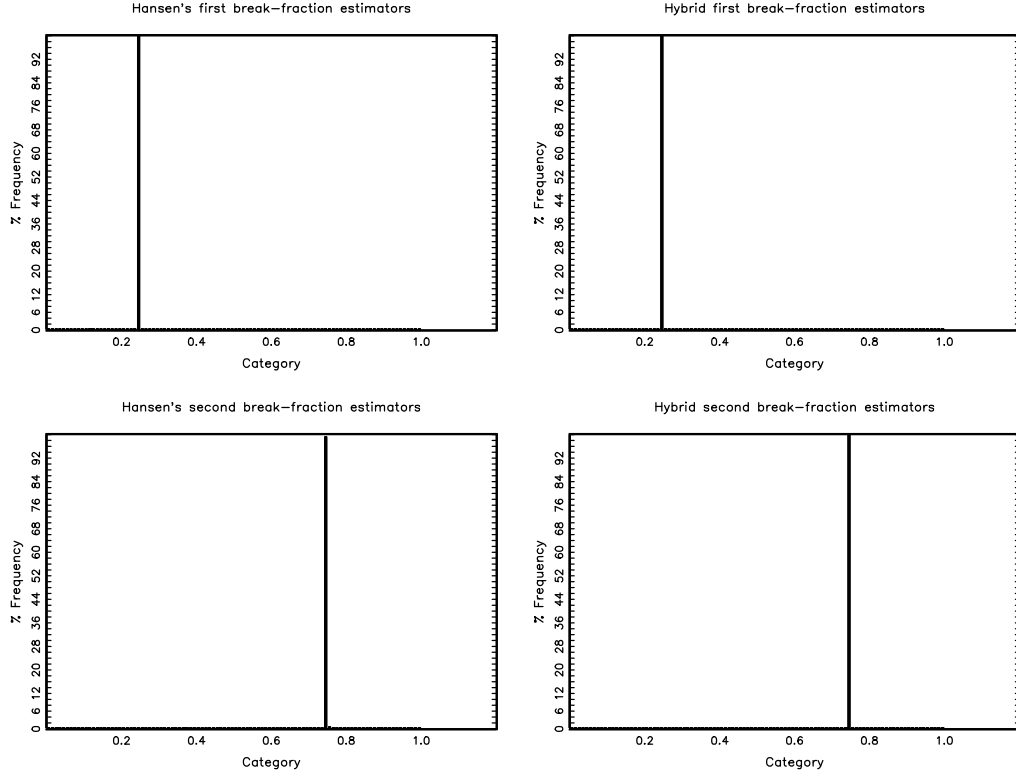


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

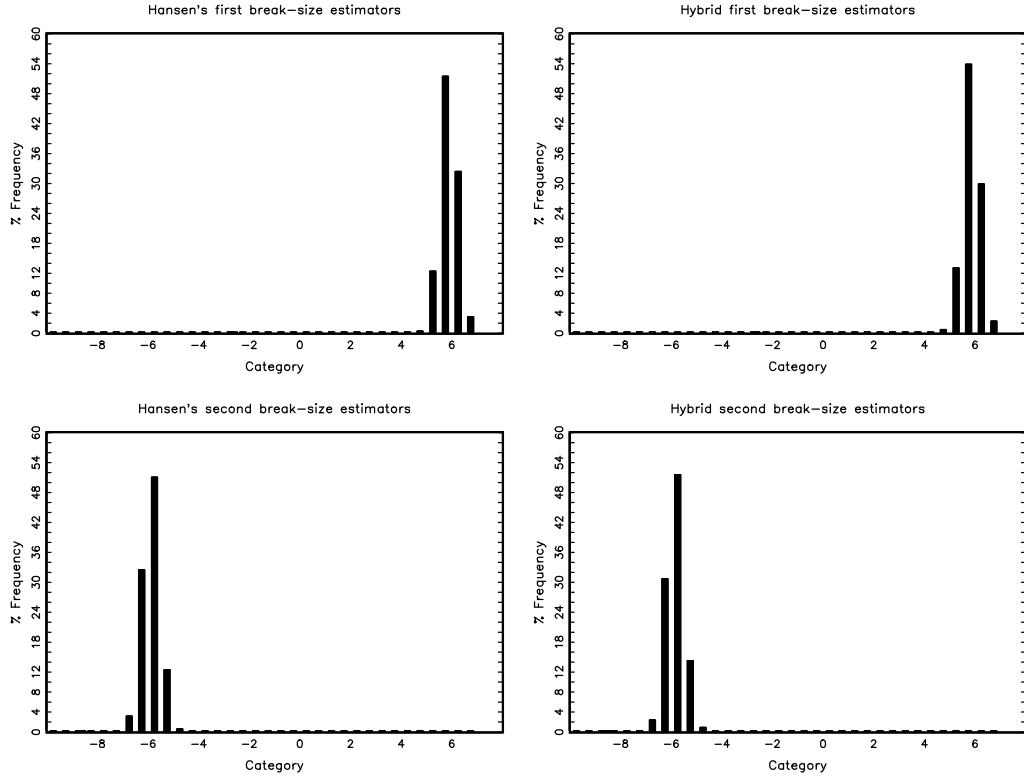


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-20: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 0.5$ ,  $\gamma_1 = -5$ ,  $\gamma_2 = 5$ .

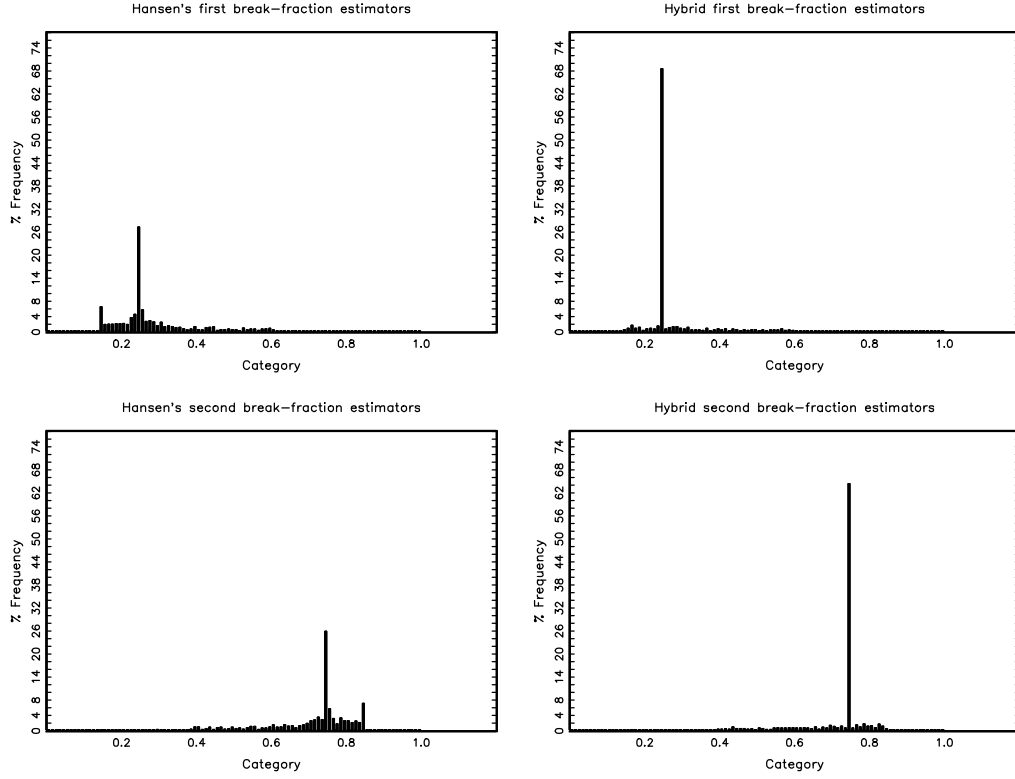


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

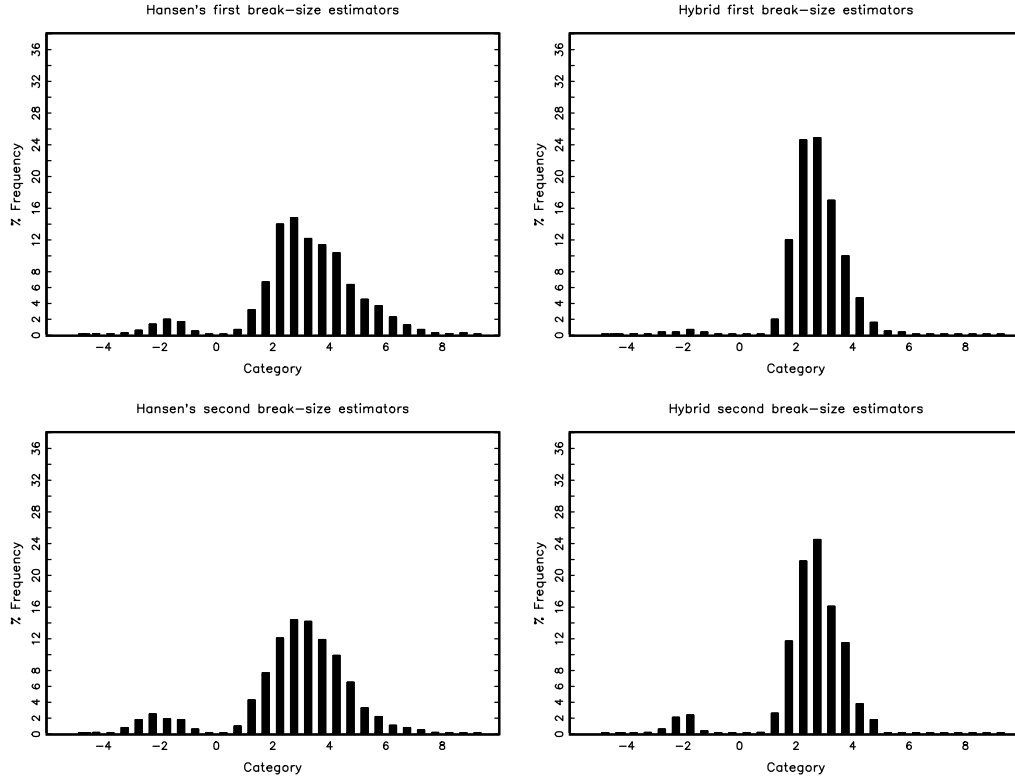


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-21: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 0.5$ ,  $\gamma_1 = 6$ ,  $\gamma_2 = -6$ .

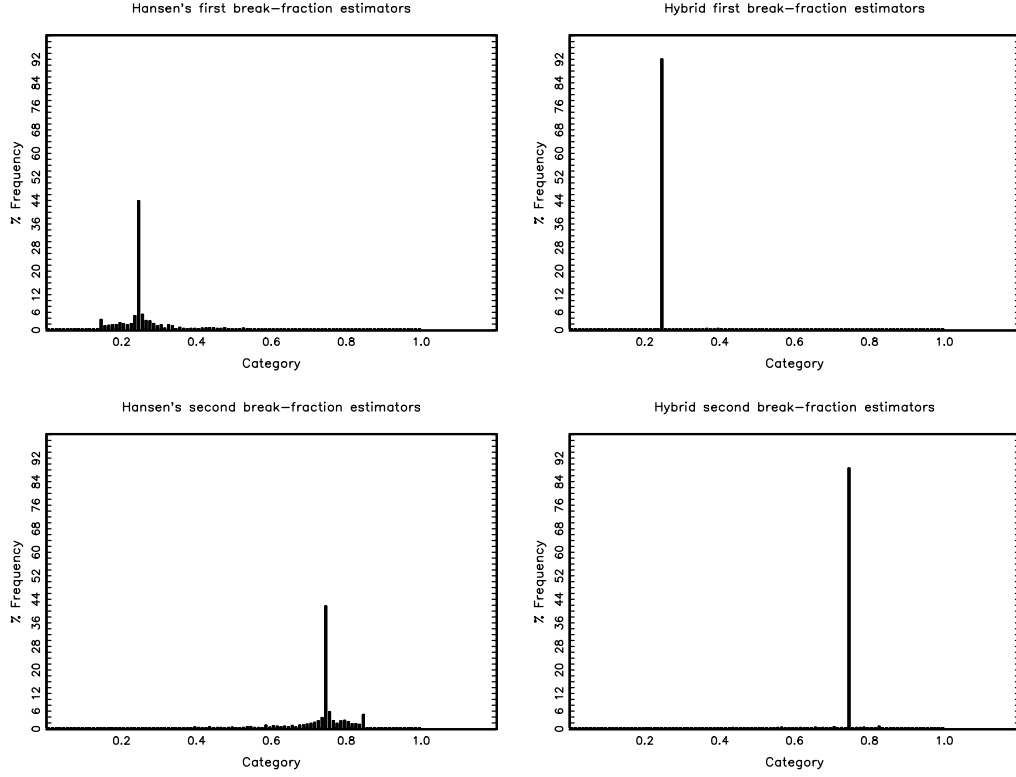


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

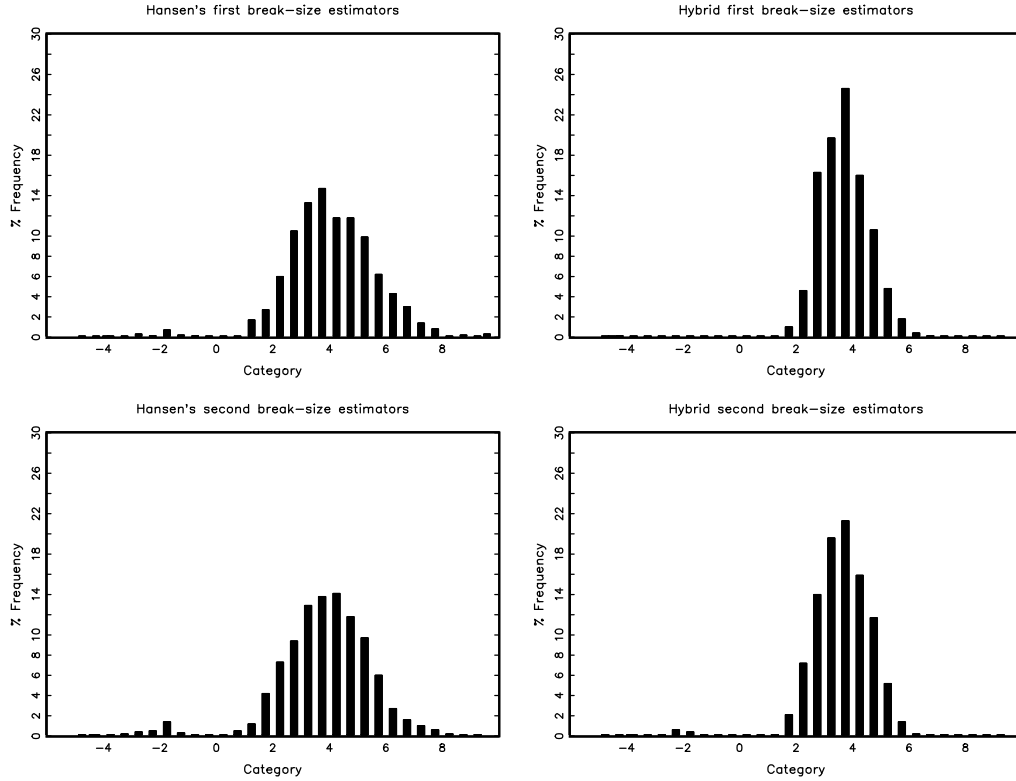


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-22: Finite sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 0.9$ ,  $\gamma_1 = 3$ ,  $\gamma_2 = 3$ .

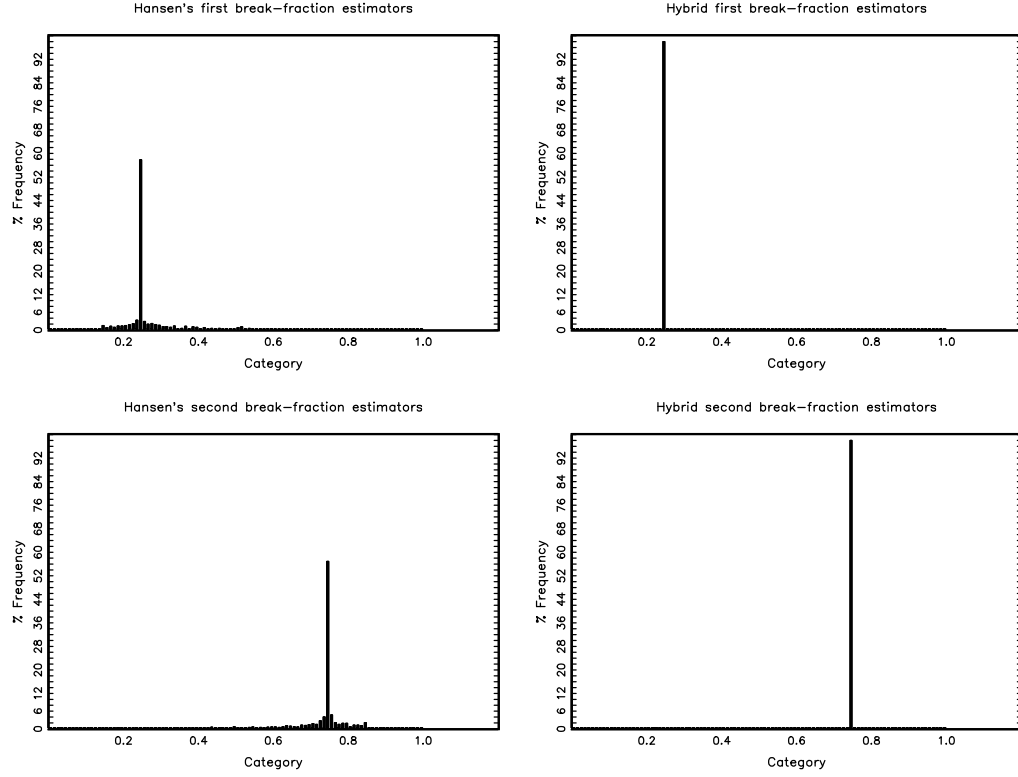


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

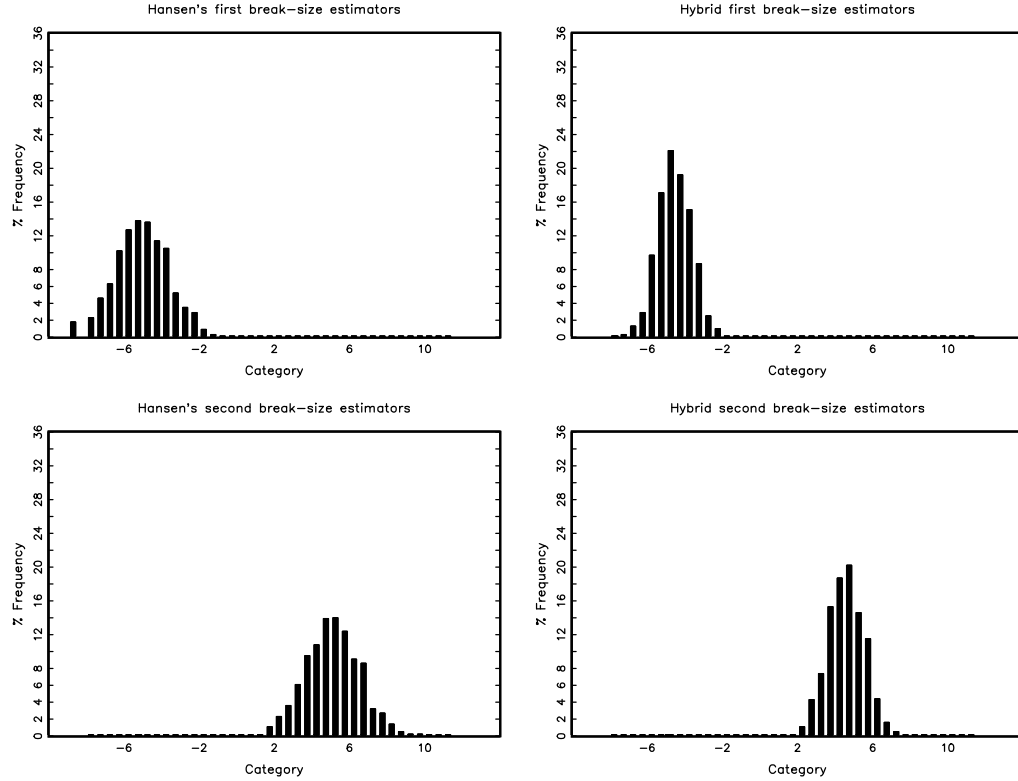


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-23: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 0.9$ ,  $\gamma_1 = 4$ ,  $\gamma_2 = 4$ .



a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

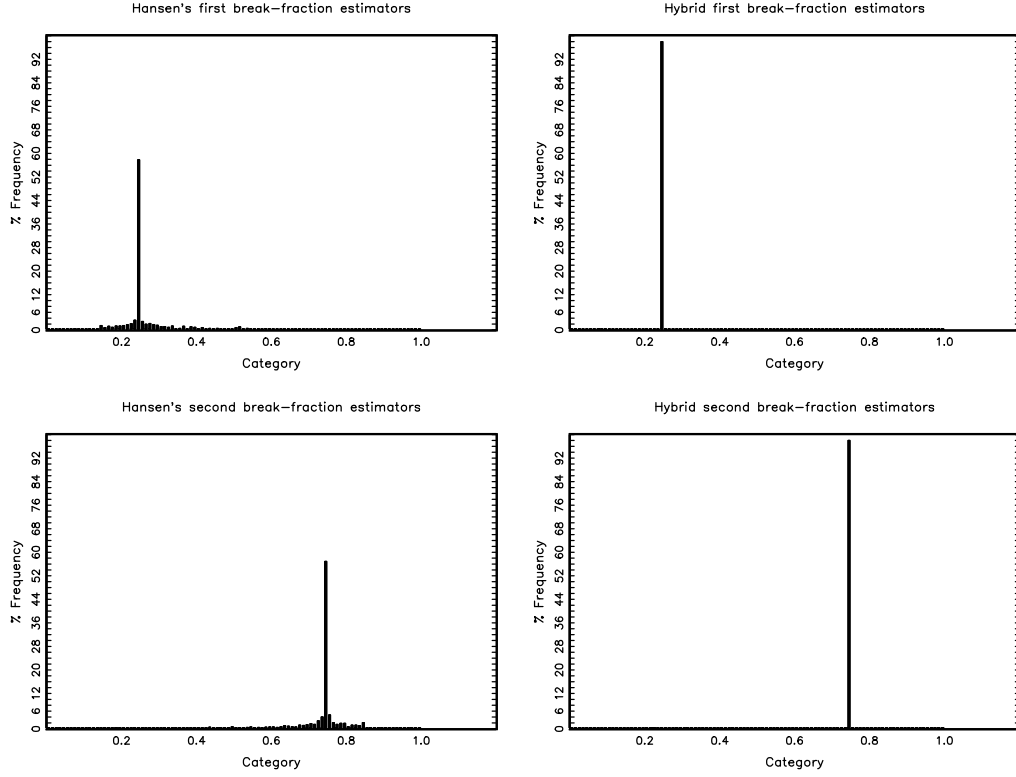


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

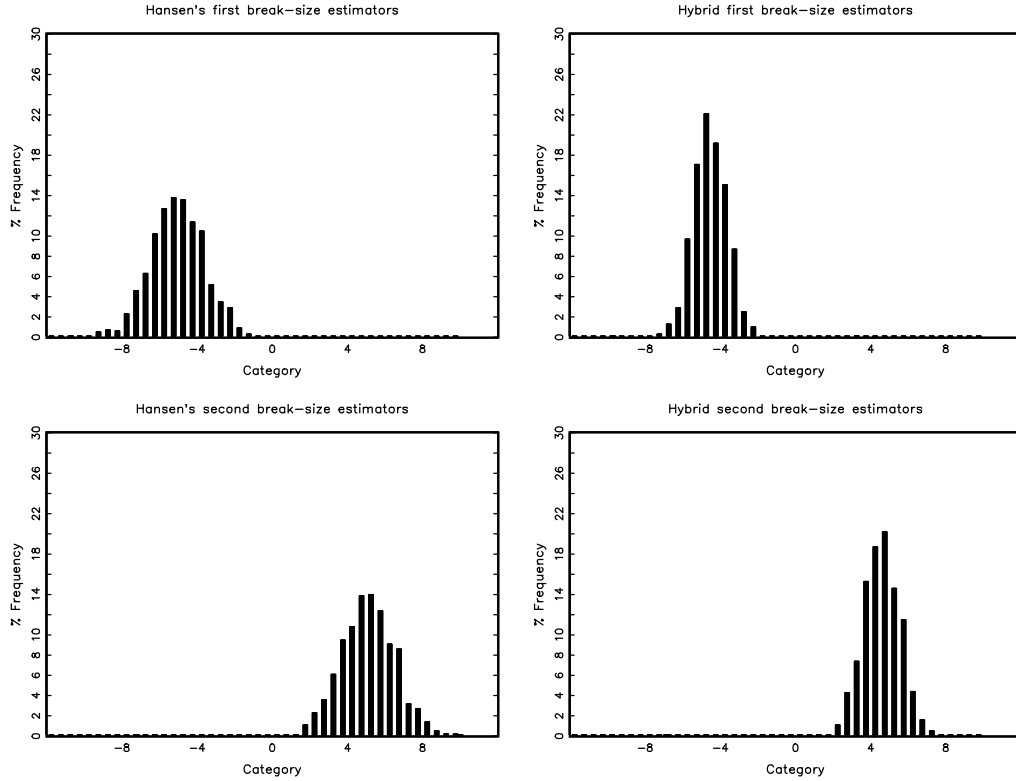
Figure 3-24: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,

$\rho = 0.9$ ,  $\gamma_1 = -5$ ,  $\gamma_2 = 5$ .



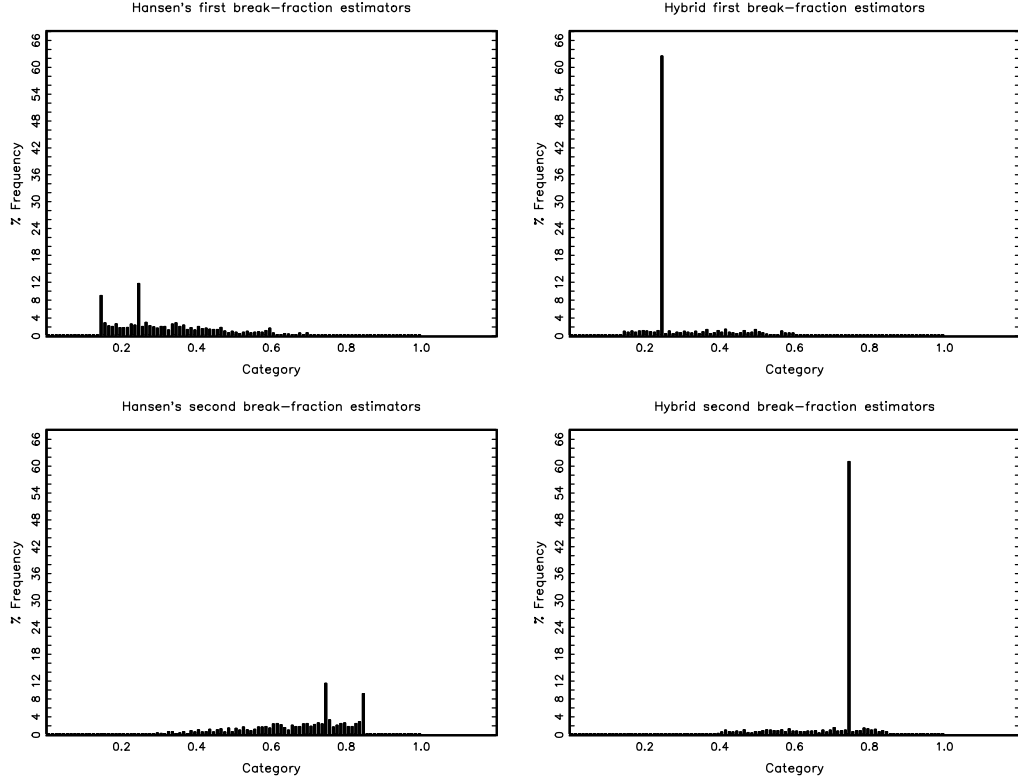


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

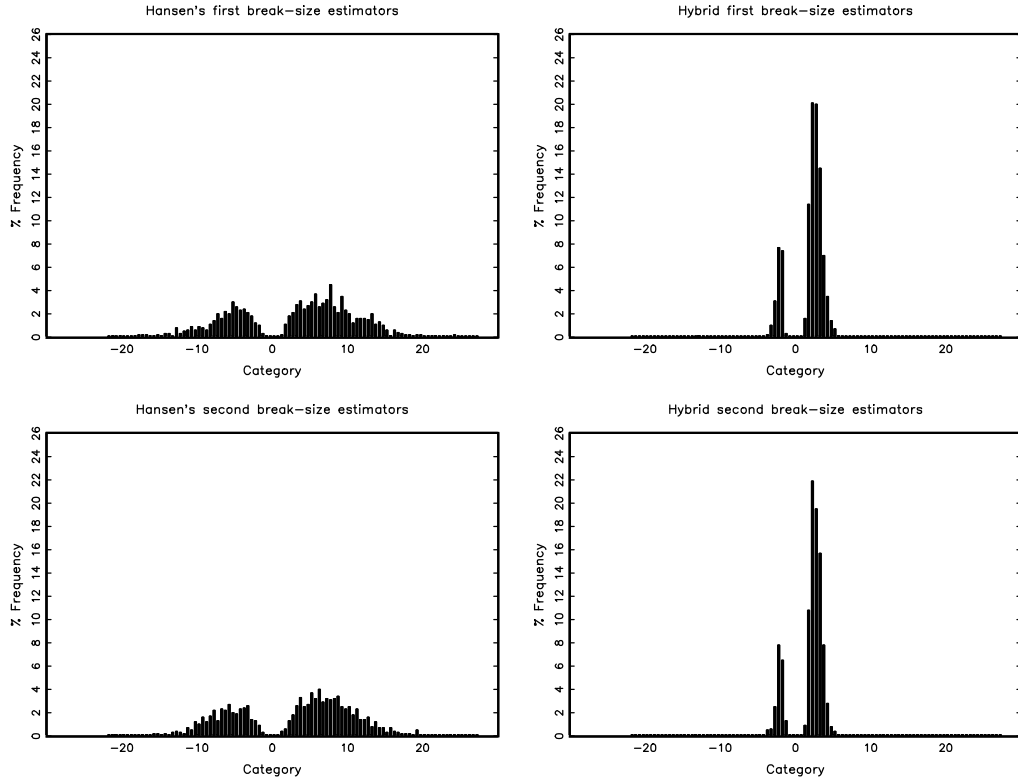


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-25: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 0.9$ ,  $\gamma_1 = 6$ ,  $\gamma_2 = -6$ .

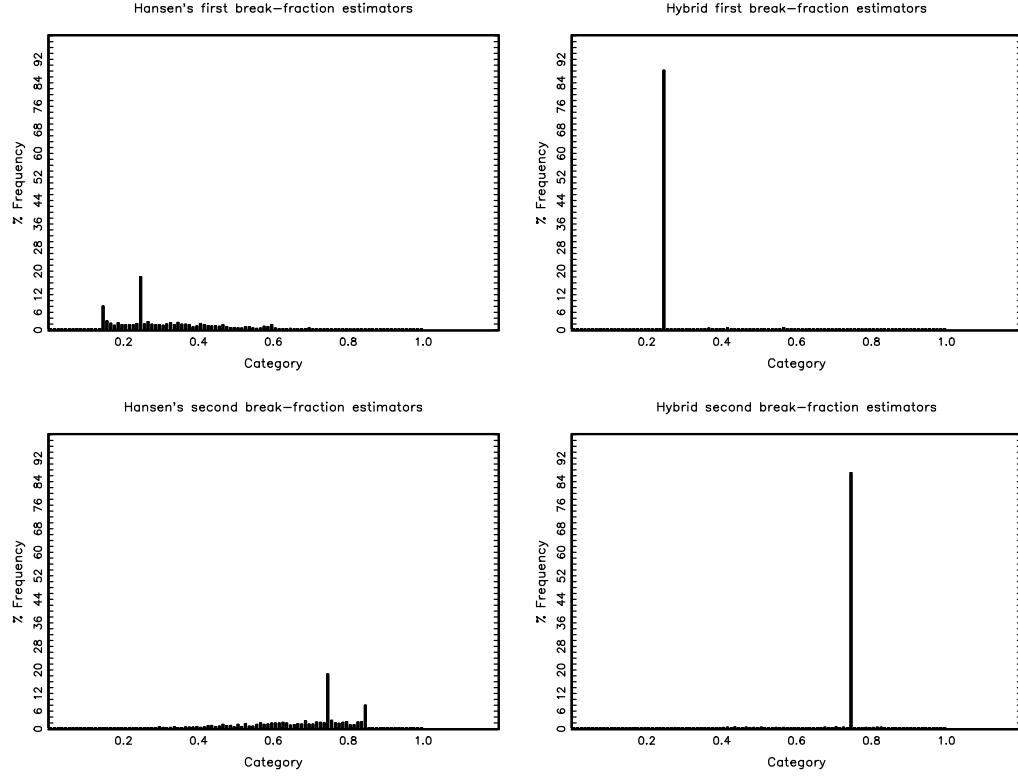


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

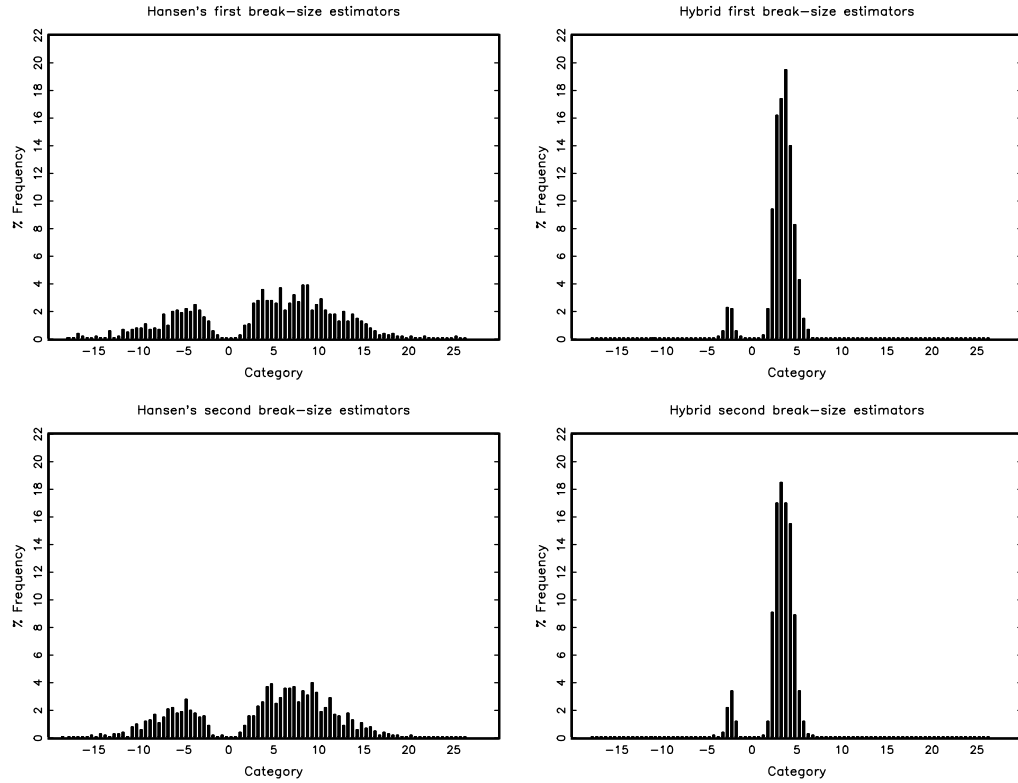


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-26: Finite sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 1$ ,  $\gamma_1 = 3$ ,  $\gamma_2 = 3$ .

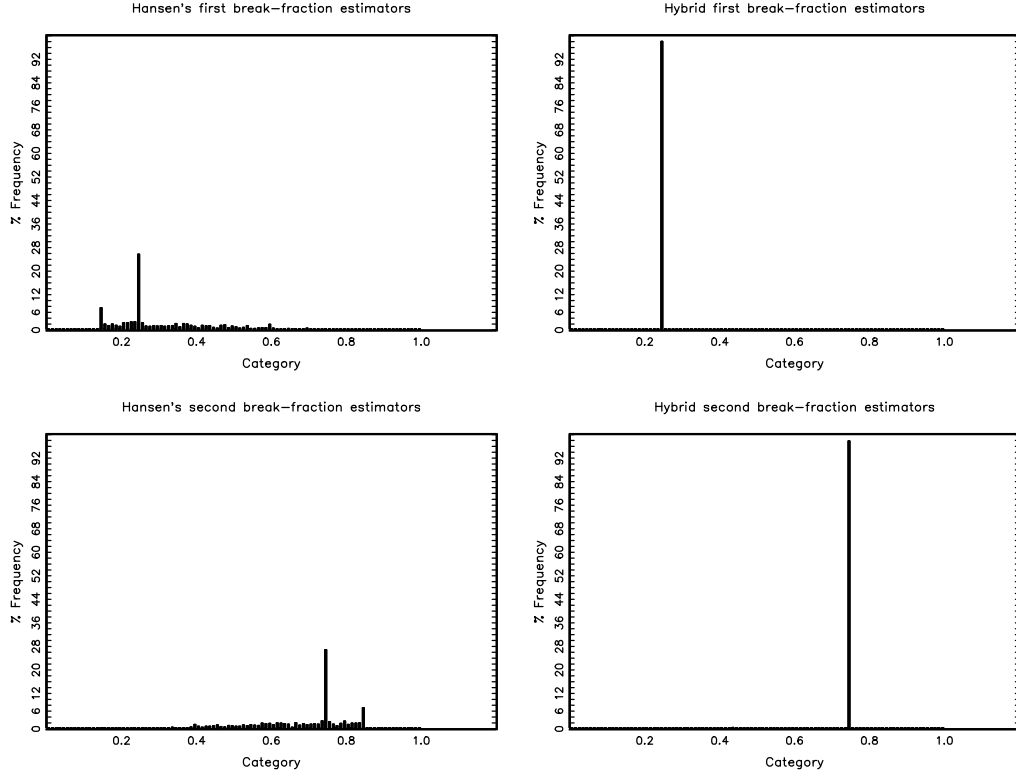


a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches

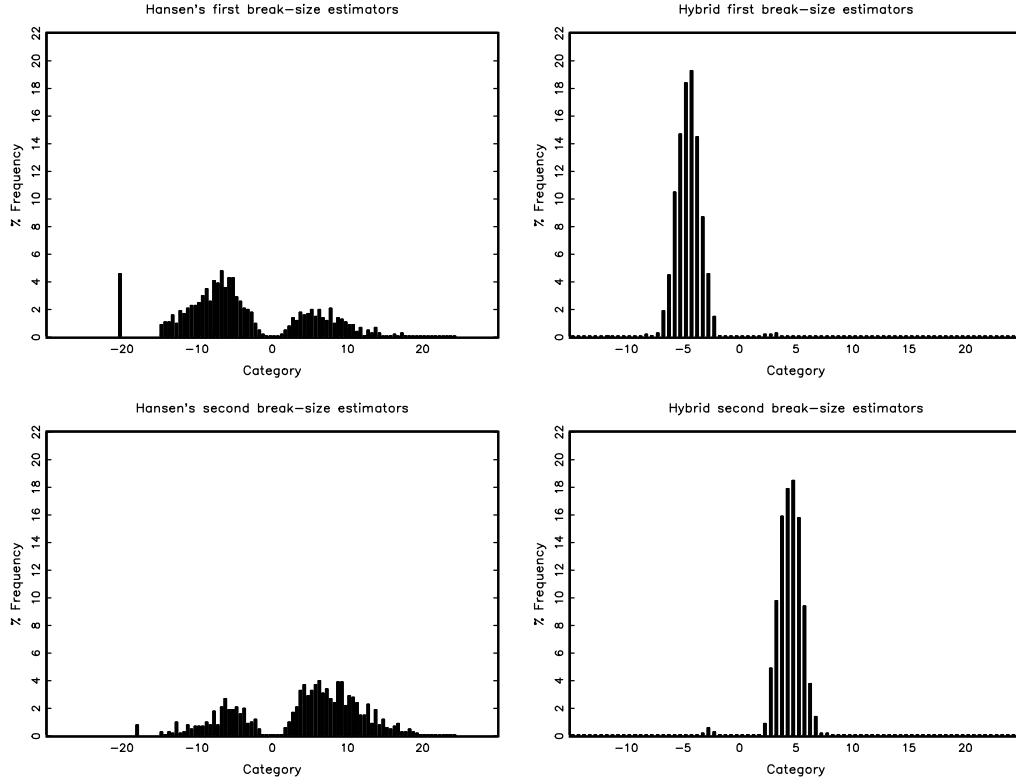


b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-27: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,  $\rho = 1$ ,  $\gamma_1 = 4$ ,  $\gamma_2 = 4$ .



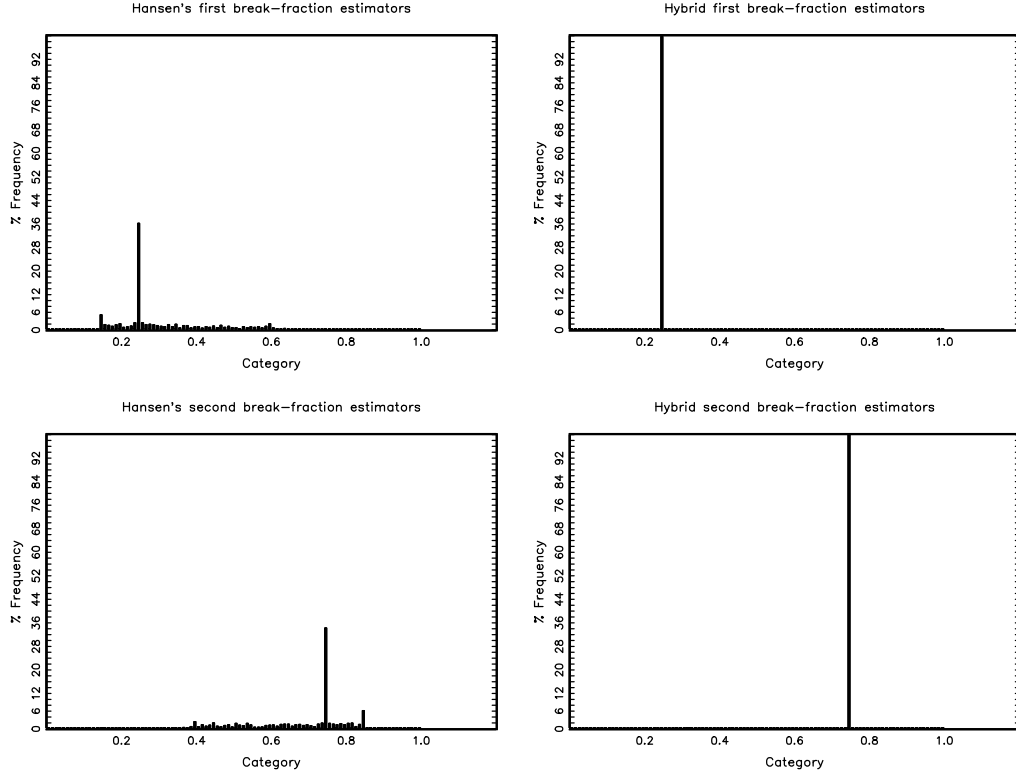
a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches



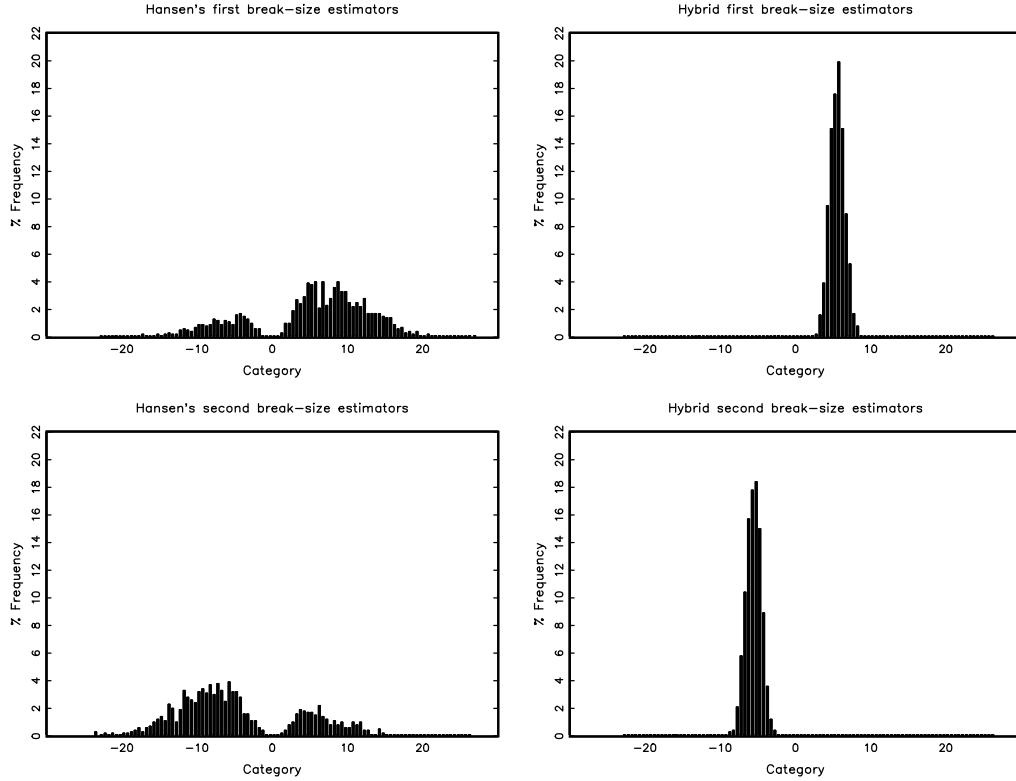
b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-28: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,

$\rho = 1$ ,  $\gamma_1 = -5$ ,  $\gamma_2 = 5$ .



a) Histograms of break-fraction estimators as given by Hansen and Hybrid approaches



b) Histograms of the two level break-size estimators as given by Hansen and Hybrid approaches

Figure 3-29: Finite Sample performance of the Hansen and Hybrid approaches with  $\tau_1^* = 0.25$ ,  $\tau_2^* = 0.75$ ,

$\rho = 1$ ,  $\gamma_1 = 6$ ,  $\gamma_2 = -6$ .

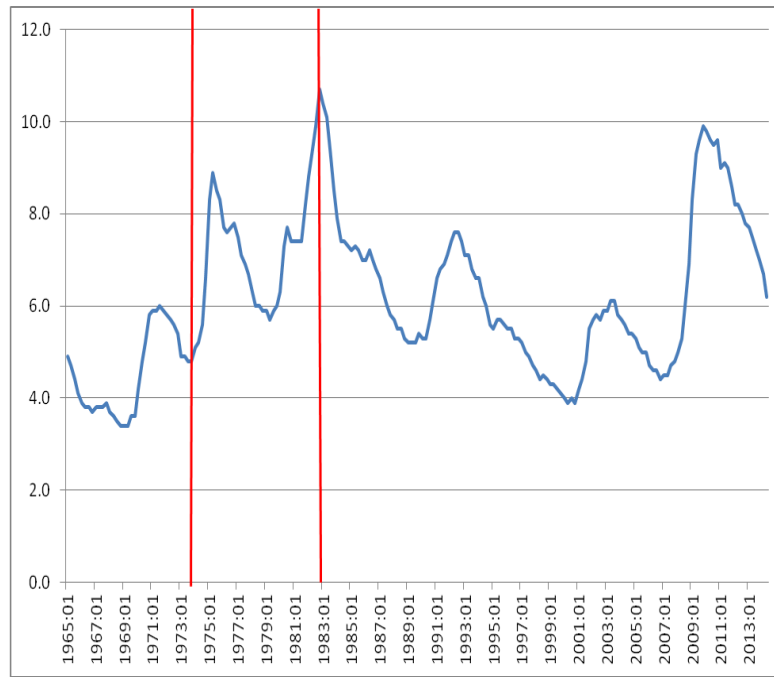


Figure 3-30: Percentage (%) of Quarterly U.S. Civilian Unemployment Rate (Seasonally Adjusted).

### 3.8 References

1. Andrews, D.W.K. (1993), "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, pp.821-856.
2. Andrews, D.W.K. and W. Ploberger (1994), "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62 (6), pp.1383-1414.
3. Andrews, D.W.K., I. Lee and W. Ploberger (1996), "Optimal change point tests for normal linear regression," *Journal of Econometrics*, 70, pp.9-38.
4. Arestis, P. and I. Biefang-Frisancho Mariscal (1999), "Unit roots and structural breaks in OECD unemployment," *Economics Letter*, 65, pp.149-156.
5. Bai, J. (1994), "Least squares estimation of a shift in linear processes," *Journal of Time Series Analysis*, 15, pp.453-472.
6. Bai, J. (1997), "Estimation of a change point in multiple regression models," *Review of Economic and Statistics*, 79, pp.551-563.

7. Bai, J. (1999), "Likelihood ratio tests for multiple structural changes," *Journal of Econometrics*, 91, pp.299-323.
8. Bai, J. and P. Perron (1998), "Estimating and testing linear models with multiple structural Changes," *Econometrica*, 66, pp.47-78.
9. Bai, J. and P. Perron (2003), "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18, pp.1-22.
10. Bai, J. and P. Perron (2006), "Multiple structural change models: a simulation analysis," In *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, eds. D. Corbea, S. Durlauf and B.E. Hansen. New York: Cambridge University Press.
11. Bai, J., R.L. Lumsdaine and J.H. Stock (1998), "Testing for and dating breaks in multivariate time series," *Review of Economic Studies*, 65, pp.395-432.
12. Blanchard, O. and L.H. Summers (1986), "Hysteresis and the European Unemployment Problem," *NBER Macroeconomics Annual*. MIT Press, Cambridge, MA.
13. Brown, R.L., J. Durbin and J.M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society, Series B*, 35, pp.149-192.
14. Camarero, M. and C. Tamarit (2004), "Hysteresis vs. natural rate of unemployment: new evidence for OECD countries," *Economics Letters*, 84, pp.413-417.
15. Camarero, M., Carrion-i-Silvestre, J.L. and C. Tamarit (2006), "Testing for hysteresis in unemployment in OECD countries. New evidence using stationarity panel tests with breaks," *Oxford Bulletin of Economics and Statistics*, 68, pp.167-182.
16. Carrion-i-Silvestre, J.L., D. Kim, and P. Perron (2009), "GLS-based unit root tests with multiple structural breaks both under the null and the alternative hypotheses," *Econometric Theory*, 25, pp.1754-1792.
17. Chong, T.T.L. (2001), "Structural change in AR(1) models," *Econometric Theory*, 17, pp.87-155.

18. Chow, G.C. (1960), "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 52, pp.211-22.
19. Clemente, J., L. Lanaspa and A. Montanes (2005), "The unemployment structure of the US states," *The Quarterly Review of Economics and Finance*, 45, pp.848-868.
20. Cochrane, J. (1991), "A critique of the application of unit root tests," *Journal of Economic Dynamics and Control*, 15, pp.275-284.
21. Davies, R.A. (1977), "Hypothesis Testing When a Nuisance Parameter is Present only Under the Alternative," *Biometrika*, 64, pp.247-254.
22. DeJong, D.N., J.C. Nankervis, N.E Savin and C.H. Whiteman (1992), "The power problems of unit root tests in time series with autoregressive errors," *Journal of Econometrics* 53, pp.323-343.
23. Elliott, G. and J. H. Stock (1994), "Inference in time series regression when the order of integration of a regressor is unknown," *Econometric Theory*, 10, pp.672-700.
24. Elmskov, J. and M. MacFarlan (1993), "Unemployment persistence". *OECD Economic Studies* 21, pp.60-89.
25. Gil-Alana, L.A. (2002), "Structural breaks and fractional integration in the US output and unemployment rate," *Economics Letters*, Vol.77, Issue 1, pp.79-84.
26. Hansen, B. E. (2009), "Averaging estimators for regressions with a possible structural break," *Econometric Theory*, 35, pp.1498-1514.
27. Hansen, B. E. (2010), "Averaging estimators for autoregressions with a near unit root," *Journal of Econometrics*, 158, pp.142-155.
28. Harris, D., D.I. Harvey, S.J. Leybourne and A.M.R. Taylor (2009), "Testing for a unit root in the presence of a possible break in trend," *Econometric Theory*, 25, pp.1545-1588.
29. Harvey, D.I. and S.J. Leybourne (2013), "Break Date Estimation for Models with Deterministic Structural Change," *Oxford Bulletin of Economics and Statistics*, 0305-9049, pp.1-20.



30. Harvey, D.I, S.J. Leybourne and A.M.R. Taylor (2009), "Simple, robust and powerful tests of the breaking trend hypothesis", *Econometric Theory*, Vol.25, pp.995-1029.
31. Harvey, D. I., S. J. Leybourne and A.M.R. Taylor (2010), "Robust methods for detecting multiple level breaks in autocorrelated times series," *Journal of Econometrics*, Vol.157, pp.342-358.
32. Harvey, D.I, S.J. Leybourne and A.M.R. Taylor (2012), "Unit root testing under a local break in trend," *Journal of Econometrics*, 167(1), pp. 140-167.
33. Kim, H.J. and D. Siegmund (1989), "The Likelihood Ratio Test for a Change-Point in Simple Linear Regression," *Biometrika*, 76, 3, pp.409-23.
34. Kuan, C.M. and K. Hornik (1995), "The Generalized Fluctuation Test: A Unifying View," *Econometric Reviews*, 14, pp.135-161.
35. Lee, C.C., C.P. Chang (2008), "Unemployment hysteresis in OECD countries: Centurial time series evidence with structural breaks," *Economic Modelling*, 25, pp.312-325.
36. Lee, J.D., C.C. Lee, C.P. Chang (2009), "Hysterestis in unemployment revisited: Evidence from panel LM unit root tests with heterogeneous structural breaks" *Bulletin of Economic Research*, 61, 4, pp. 325-334.
37. Mallows, C.L. (1973), "Some comments on Cp". *Technometrics* 15, pp.661-675.
38. Mitchell, W.F. (1993), "Testing for unit roots and persistence in OECD unemployment rates," *Applied Econometrics*, 25, pp.1489-1501.
39. Papell, D.H., C.J. Murray and H. Ghiblawi (2000), "The structure of unemployment," *The Review of Economics and Statistics* 82, pp.309-315.
40. Perron, P. (1989), "The great crash, the oil price shock and the unit root hypothesis," *Econometrica* 57, pp.1361-1401.
41. Perron, P. and T. Yabu (2009), "Testing for shifts in trend with an integrated or stationary noise component," *Journal of Business and Economic Statistics*, Vol.27, pp.369-396.

42. Perron, P. and X. Zhu (2005), "Structural breaks with deterministic and stochastic trends," *Journal of Econometrics*, Vol. 129, pp.65-119.
43. Perron, P. and Z. Qu (2006), "Estimating Restricted Structural Change Models," *Journal of Econometrics*, 1, 34, pp.373-399.
44. Pesaran, M.H., D. Pettenuzzo and A. Timmermann (2006), "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, Wiley Blackwell, vol. 73(4), pp.1057-1084.
45. Ploberger, W. and W. Krämer (1990), "The local power of the cusum and cusum of squares tests," *Econometric Theory* 6, pp.335-347.
46. Ploberger, W., W. Krämer and K. Kontrus (1989), "A new test for structural stability in the linear regression model," *Journal of Econometrics*, 40, pp.307-318.
47. Quandt, R.E. (1960), "Tests of the hypothesis that a linear regression system obeys two separate regimes," *Journal of the American Statistical Association*, 55, pp.324-330.
48. Stock, J.H. and M.W. Watson (1996), "Evidence on structural instability in macroeconomic time series relations", *Journal of Business and Economic Statistics*, Vol. 14, pp.11-30.
49. Stock, J.H. and M.W. Watson (1999), "A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series", in Engle R. F. and White H. (eds), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, Oxford University Press, Oxford, pp. 1-44.
50. Stock, J.H. and M.W. Watson (2005), "Implications of Dynamic Factor Analysis for VAR Models," NBER Working Paper No. 11467.
51. Vogelsang, T.J. (1998), "Testing for a shift in mean without having to estimate serial-correlation parameters", *Journal of Business and Economic Statistics*, Vol.16, pp.73-80.

# Chapter 4

## Regime Change, Inflation Dynamics and the meta-Phillips Curve

### 4.1 Introduction

There has been a long history of examining the inflation process empirically through estimated Phillips curves, culminating in the New Keynesian Phillips Curve (NKPC) which is now widely adopted in macroeconomic models. Despite plenty of evidence of shifts in the mean and persistence of inflation over time in many countries, many studies based on the NKPC assume that structurally stable relationships exist. This means different inflationary episodes might be misinterpreted and tests of the underlying behavioural relations could be misleading. For example, Russell et al. (2010) demonstrate empirically that Galí and Gertler's (1999) widely-believed view that forward-looking inflation expectations are important in determining inflation today could be due to a failure to adequately take into account structural breaks. The numerous shifts in the level of inflation, volatility changes, the decreasing slope of the Phillips Curve and the decline in inflation persistence are just some of the dramatic changes that took place in inflation dynamics that make modelling and forecasting inflation a challenging task.

In this chapter, we describe an analysis of the U.S. inflation which is able to accommodate structural instability arising from regime breaks and changes in the underlying drivers of price-setting decisions, in a flexible way. The analysis follows the approach of Lee, Morley and Shields (2015) who suggest combining Taylor rule models that are estimated over different sample periods using model averaging techniques. Our work contributes to literature by constructing a meta-Phillips Curve, which involves estimation and inference of a set of specific Phillips Curves obtained

through Generalised Methods of Moments (GMM), but combined using Bayesian Model Averaging (BMA) techniques. The weights employed in combining individual Phillips Curves to obtain the "meta-Phillips Curve" are determined according to the ability of the individual Phillips Curves to explain past inflation behaviour. The fact that weights change over time provides a useful and flexible structure with which we can interpret the changing inflation dynamics. The analysis shows that, despite the considerable structural instability observed, the meta-Phillips Curve provides a useful vehicle with which to explain inflation dynamics, and supports the view that forward-looking expectations play a key role in inflation determination. The estimated meta-Phillips Curve also provides a coherent characterisation of inflation dynamics in the U.S. over the last fifty years, often matching regime changes in monetary policy and Central Bank's reactions to economic situations.

The remainder of this chapter is as follows: Section 4.2 discusses recent developments in the Phillips Curve literature and discusses the considerable structural uncertainty embedded in inflation dynamics. Section 4.3 describes the modelling approach, focusing on the estimation method and the construction of weights. Section 4.4 presents the results of the estimation of the U.S. meta-Phillips Curve over the period 1959 :  $Q4$  – 2012 :  $Q4$ , emphasising the phases of inflation dynamics in which expectations were more or less anchored, where anti-inflationary policies were pursued more or less aggressively and when responses to the real economic activity became more or less acute. Section 4.5 concludes.

## 4.2 Literature Review

### 4.2.1 The New Keynesian Phillips Curve

The collapse of the traditional Phillips Curve gave way to the expectations-augmented Phillips Curve:

$$\pi_t = a + \gamma U_t + \rho \sum_{i=1}^N \beta_i \pi_{t-i} + \varepsilon_t,$$

where  $\pi_t$  and  $U_t$  respectively denote the inflation and unemployment rate at time  $t$ , assuming that agents form expectations adaptively. The expectations-augmented Phillips Curve unfolded a renowned sequence of heavily cited research investigating inflation dynamics using the unem-

ployment rate as the forcing variable. The NKPC, on the other hand, emerged to reconcile the microfoundations of the Real Business Cycle School with rational forward-looking expectations. The NKPC specification abandons the trade-off between unemployment and inflation and introduces a markup gap as the forcing variable for inflation dynamics. The NKPC is, then, given by:

$$\pi_t = \beta E_t \pi_{t+1} + \gamma mc_t, \quad (4.1)$$

where  $E_t \pi_{t+1}$  represents inflation expectations for  $t + 1$  as formed at  $t$ , and  $mc_t$  denotes real marginal cost. The NKPC in (4.1) is derived from Rotemberg's (1982) and Calvo's (1983) price setting formulation as a solution to firms' profit maximisation problem. These models are set in a monopolistically competitive environment, in which firms are constrained by the frequency at which they can adjust their prices, for example due to menu costs, and they all reduce to the forward-looking Phillips Curve as shown by Roberts (1995) and Clarida, Galí and Gertler (1999).

Note that under some assumptions about the labour supply process (Rotemberg and Woodford (1997)), the output gap is linearly related to the real marginal cost, implying that  $mc_t = kx_t$ , where  $x_t$  is the difference between the log of output and the log of its natural level i.e.  $x_t = y_t - y_t^*$  so that equation (4.1) can be written as:

$$\pi_t = \xi x_t + \beta E_t \{\pi_{t+1}\} \quad (4.2)$$

Iterating forward equation (4.2) gives

$$\pi_t = \xi k \sum_{k=0}^{\infty} \beta^k E_t \{x_{t+k}\}, \quad (4.3)$$

implying that inflation depends exclusively on the discounted sequence of future output gaps. Despite its popular implementation, the model met fierce criticisms about the extent it really reflects the observed inflation behaviour. In particular, the New Keynesian Phillips curve in equation (4.3) implies that disinflation of any size can be achieved costlessly and immediately by a central bank that could commit to setting the path of future output gaps to zero, something which is not supported empirically. In addition, Fuhrer (1997) and Roberts (1998) provide evidence that inflation dynamics are characterised by a high degree of persistence which the NKPC in (4.2) fails to account for. Although the NKPC is based on price level stickiness, the inflation is assumed to be perfectly flexible something which is also out-of-line with the empirical evidence.

### 4.2.2 The Hybrid New Keynesian Phillips Curve

The well-argued weaknesses of the New Keynesian Phillips Curve led to the emergence of the Hybrid Phillips Curve as a convex combination of expected future inflation and lagged inflation. The addition of the lagged inflation serves to capture the observed inflation persistence while implying that disinflations now involve costly output reductions. Several theoretical motivations led to the Hybrid Phillips Curve (e.g. the relative wage model by Fuhrer and Moore (1995 a, b), indexation by Christiano et al. (2005)) but here we concentrate on the rule of thumb price-setting behaviour as captured by Galí and Gertler (1999).

#### Rule of thumb price-setters

Here we follow Galí and Gertler (1999) who extend the basic Calvo model and capture inflation inertia by allowing a subset of firms to use a backward-looking rule of thumb when setting their prices. In particular, following Calvo (1983), the probability that each firm is able to change its price in any given period is  $1 - \theta$ . From those firms that adjust their prices, a proportion  $1 - \omega$  are forward looking and set their prices optimally taking into account the constraints on the timing of the adjustments using all available information to forecast future marginal costs. The remaining fraction of firms,  $\omega$ , are backward looking using a rule of thumb based on the history of price behaviour. This allows us to write the aggregate price level as:

$$p_t = \theta p_{t-1} + (1 - \theta) p_t^*,$$

where  $p_t^*$  is the price newly set in period  $t$  which itself evolves according to:

$$p_t^* = (1 - \omega) p_t^f + \omega p_t^b,$$

where,  $p_t^f$  is the price set by forward looking firms and  $p_t^b$  is the price set by backward-looking firm. Because the forward looking firms behave exactly in the same manner as in the Calvo setting,  $p_t^f$  can be expressed as:

$$p_t^f = (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k E_t \{ mc_{t+k}^n \}.$$

Assuming that backward-looking price setters rely solely on information in period  $t-1$  or earlier to adjust their price in period  $t$ , and that in the steady state the rule is consistent with optimal

behaviour,  $p_t^b$  can be written as:

$$p_t^b = p_{t-1}'^* + \pi_{t-1},$$

where  $p_{t-1}'^*$  denotes the average price set in period  $t - 1$ . This indicates that backward looking firms set their prices based on past price adjustments corrected by previous inflation. Although ad hoc, this framework allows the rule to converge to the optimal price in the long-term provided that inflation is a stationary process. Moreover, because  $p_{t-1}'^*$  partially incorporates the forward-looking behaviour, the deviation of firms that use rule of thumb from the optimal price will be of second order, given that the percent difference between backward and forward price is not large. As Galí and Gertler (1999) emphasise, this is likely to be the case when the fraction of backward-looking firms is small. Putting all the above equations together yields the hybrid Phillips Curve:

$$\pi_t = \lambda mc_t + \gamma_f E_t\{\pi_{t+1}\} + \gamma_b \pi_{t-1}, \quad (4.4)$$

where

$$\begin{aligned} \lambda &= (1 - \omega)(1 - \theta)(1 - \beta\theta)\phi^{-1}, \\ \gamma_f &= \beta\theta\phi^{-1} \\ \gamma_b &= \omega\phi^{-1} \\ \phi &= \theta + \omega[1 - \theta(1 - \beta)]. \end{aligned}$$

The derivation of the hybrid Phillips Curve allows the coefficients to be functions of deep model parameter:  $\theta$ , which measures the degree of price stickiness;  $\omega$ , which reflects the fraction of backward-looking price setters; and  $\beta$ , the discount factor. One weakness of this rationale behind the hybrid model is that it is unclear why some parts of the population appear to be less sophisticated than others.

### 4.2.3 Issues in empirical application

The hybrid model sparked off an ongoing debate over its empirical performance. The importance of empirical evidence on the forward- and backward-looking components in the hybrid Phillips Curve models is crucial for policy prescriptions. Although many studies are based on a common dataset, findings are mixed due to the choice of the forcing variable, estimation methods, specification issues

and choice of the instrument set. In this subsection we summarise the most important problems faced in empirical work.

### **The choice of the forcing variable**

Galí and Gertler (1999) and Galí and Gertler and Lopez-Salido (2005) employ GMM using unit labour cost as the measure of marginal cost, and find that inflation inertia is a much less significant contributor to current inflation than what Fuhrer and Moore (1995 a, b) and Fuhrer (1997) claim it to be, concluding that the purely forward-looking NKPC provides a good approximation to the dynamics of inflation. Consistent with Galí and Gertler (1999), Sbordone (2002, 2005) uses a two-step estimation procedure to confirm the dominant role of the forward-looking term in the hybrid Phillips Curve. As Roberts (2001) emphasises, the conflicting evidence on the performance of the Hybrid NKPC can be rationalised by the choice of the forcing variable. Gagnon and Khan (2005) highlight that the choice of the production function and the decision about the relevant measure of marginal cost dramatically influence the empirical weight on the backward-looking term.

### **Omitted variable bias**

Rudd and Whelan (2005b) demonstrate that the omitted variable bias and specification issues raise questions over the empirical validity of results based on the NKPC. As Rudd and Whelan emphasise, the omitted variable problem means that the regression error is no longer an expectational error since it also includes the effect of the omitted variables and, therefore, estimates on the forward-looking component will be biased upwards as long as future inflation and its instruments are correlated with the omitted variable. As Rudd and Whelan (2006, p.13) indicate, "Galí and Gertler included additional lags of inflation, commodity prices, and detrended output in their instrument set" and consequently "the constructed proxy for  $E_t\pi_{t+1}$  will capture the influence of these omitted variables and receive a large coefficient even if  $E_t\pi_{t+1}$  itself has no independent influence whatsoever on inflation."



## The weak instruments problem

The parameters of the hybrid NKPC are typically estimated by replacing the unobserved term  $E_t(\pi_{t+1})$  in equation (4.4) by  $\pi_{t+1} - \eta_{t+1}$ , where  $\eta_{t+1}$  is the one-step-ahead forecast error in  $\pi_{t+1}$  so that the transformed equation can be estimated using any predetermined variables as instruments. Instruments are weak whenever the correlation of the endogenous variable with the instruments is small relative to the sample size. The weak identification leads to GMM statistics with non-normal distributions, so that the conventional GMM inferences are misleading. As Atkeson and Ohanian (2001) emphasise, the NKPC is weakly identified because changes in inflation are hard to predict so that instruments which are thought to be exogenous are almost irrelevant. Mavroeidis (2005) shows that identification failure is likely to occur when the forcing variable is irrelevant for inflation determination such that the Phillips Curve is almost flat. When cost-push shocks are unpredictable, inflation is also unpredictable so that the forward-looking term is completely unidentified since no predetermined instruments exist. Given that many studies document a flat Phillips Curve, identification failure prevails without too much controversy. In addition, Kleibergen and Mavroedis (2009) demonstrate that "the size of the identification robust statistics becomes sensitive to the number of instruments when a Heteroskedasticity and Autocorrelation Consistent (HAC) estimator is used and the size distortions can be rather large." Pesaran (1987), Ma (2002) and Mavroeidis (2005) emphasise that weak instruments may provide a convincing explanation of the conflicting evidence on the estimates of the hybrid NKPC<sup>1</sup>.

### 4.2.4 Regime Changes and the Phillips Curve

With the deep structural changes the U.S. economy has undergone through (Willis 2003) and the changing monetary policy (Judd and Rudebusch (1998) and Clarida et al. (1999, 2000)), it is very likely that inflation dynamics have experienced major shifts<sup>2</sup>. Besides, the Phillips Curve is an

---

<sup>1</sup>Dufour (1997), Staiger and Stock (1997), Stock, Wright and Yogo (2002) and Kleibergen (2002) demonstrate that standard asymptotic procedures which do not correct for local-almost-identification lead to spurious over-rejections and this is likely to be the reason why studies that rely on weak instruments yield misleading estimates.

<sup>2</sup>Inter alia, Alogoskoufis and Smith (1991), Kim and Nelson (1999) and Blanchard and Galí (2007) all provide evidence of Phillips Curve instability.

important ingredient in monetary policy analysis (Clarida, Gali and Gertler (1999)) and, thus, it is only natural that shifts in monetary policy regimes will induce changes in inflation dynamics. Because  $\gamma_b$  and  $\gamma_f$  are functions of deep parameters related to firms' pricing behaviour which is itself affected by monetary policy changes, the relative significance of these coefficients is likely to change over time.

Work by Russell et al. (2010) demonstrates that not taking into account the breaks in inflation dynamics may lead to biased and spurious estimates of the Phillips Curve. The paper establishes nine "inflation regimes" of shifting mean of U.S. inflation emphasising that the common finding of the dominance of the forward dynamic inflation term in the new-Keynesian Phillips Curve is partly attributed to the unaccounted shifts in the mean of inflation. In effect, breaks manifest themselves through the increasing importance of the forward looking term, highlighting the need to properly accommodate structural changes in order to make the correct inferences on the estimated coefficients on the dynamic terms. As the debate on the dominance of the backward-looking vis-a-vis the forward-looking term is still ongoing, accounting for breaks constitutes an important contribution to literature. We will argue that according to the prevailing economic environment and monetary policy in place, the relative importance of each term may vary. In what follows, we describe how different monetary policy regimes, as well as the prevailing economic conditions, can lead to changes in the underlying deep parameters characterising the hybrid NKPC.

- Although rational expectations has been an underlying assumption so far, we argue that according to the prevailing economic conditions and the monetary policy in place, agents may temporarily depart from rational expectations and adopt to an imperfect information environment. For instance, if information is costly, agents may wilfully decide not to obtain such information and this may be a rational choice. Similarly, if signal extraction is impeded and information becomes available with lags, forecast errors may persist due to information rigidities. We argue that private sector's learning about the monetary policy inflation target depends on the insistence of Central Bank to communicate its commitment in situations where information may become sticky. Consequently, there may be situations where expectations can become anchored or de-anchored as emphasised by Orphanides and Williams (2006, 2012), influencing the firm's price setting behaviour.

The commitment of Central Bank to maintain price stability and its strong anti-inflation stance can substantially influence the price-setting behaviour of firms. For instance, Volcker-Greenspan's adoption of a proactive stance towards managing inflation has led to a greater control over inflation expectations<sup>3</sup>. As Mishkin (2007) emphasises, with expectations of inflation well-anchored, any given shock has a more transient and smaller effect so that agents are more capable of predicting the future outcomes of the variables of interest. Accordingly, monetary policy that brings stability and assurance of the future prospects of the economy can lead to a decrease in the deep parameter  $\omega$  and an increase in  $\gamma_f$ , a decrease in  $\gamma_b$  and an increase in  $\lambda$ .

- The price-setting behaviour of firms is also affected by the nature of current inflation, as developed by the monetary policy in place. Mishkin (2007) highlights that low and less-variable inflation influences the frequency with which firms change their prices. Along the same lines, Ball, Mankiw and Romer (1988) argue that, the lower and more stable inflation regime established post-1982 led to less-frequent price adjustments. The idea is that low-inflation allows firms to leave their prices fixed for long periods of time at little cost. Accordingly, low inflation leads to an increase in the deep parameter  $\theta$  and an increase in  $\gamma_f$ , a decrease in  $\gamma_b$  and a flattening Phillips Curve, as captured by a smaller  $\lambda$  parameter.
- Inflation persistence, defined as the speed with which inflation returns to baseline after shock, is also shown to influence the price-setting behaviour of firms as documented by Taylor (2000) and Kim and Kim (2008). In particular, when inflation is highly persistent, past inflation contains more relevant information for firms' pricing practices, so that the fraction of backward-looking firms is high. The increase in  $\omega$  due to an increase in persistence can therefore lead to a decrease in  $\gamma_f$ , an increase in  $\gamma_b$  and a fall in  $\lambda$ . The decline in inflation persistence post-1980 (Cogley and Sargent (2001), Willis (2003) and Levin and Piger (2004)) was attributed to the underlying monetary policy as documented by Gaspar, Smets and Vestin (2006) validating the idea that monetary policy regime can induce structural breaks in inflation dynamics<sup>4</sup>.

---

<sup>3</sup>See Erceg and Levin (2003), Taylor (2000) and Kim and Kim (2008).

<sup>4</sup>Indeed, Roberts (2006), Borio and Filardo (2007) and Musso et al.(2009) attribute the recent flattening of the

- Globalisation and the decline in the degree to which firms “pass through” changes in costs to prices (often known as a reduction in the pricing power of firms) can also influence the deep parameters and alter parameters in the hybrid NKPC (see for example Blanchard and Galí (2007) and Taylor (2000)). Willis (2003) explains that as trade barriers decline, the increase in global competition dampens the ability of firms to increase prices so that the proportion of firms that leave prices unchanged i.e. the deep parameter  $\theta$ , increases resulting to an increase in  $\gamma_f$ , and a fall in  $\gamma_b$  and  $\lambda$ . Indeed, the well documented flattening of the Phillips Curve in recent years is attributed to the globalisation process (Ihrig et al. (2007) and Melick and Galati (2006)).

In addition to the changes in the parameters of the hybrid NKPC, monetary policy and the prevailing economic conditions can induce shifts in the drivers of inflation and the appropriate forcing variables that should enter the Phillips Curve relationship. We argue that another form of structural instability is that the best measure of resource utilisation may change over time depending on the underlying economic environment. For instance labour share (defined as the difference between the log of compensation to employees and the log of nominal GDP in deviation from sample average) may become more useful in driving the inflation process when a large proportion of firms are backward-looking since labour share is a more immediate and real-time measure of marginal cost. Conversely, the output gap (defined as the deviation of real output from its trend) is a little more forward-looking as its construction is based on both past and future values of output so that it is a more appropriate measure of resource utilisation in periods where agents are more forward-looking. Given that the debate about the appropriate measure of resource utilisation is still ongoing<sup>5</sup>, uncovering when each measure is more relevant is deemed important.

---

Phillips Curve to the monetary policy in place.

<sup>5</sup>Galí and Gertler (1999), Galí et al. (2001) and Sbordone (2002) are among the advocates of labour share while many authors such as Rudd and Whelan (2007) and Neiss and Nelson (2002) condemn using labour share as the forcing variable and strongly encourage researchers to use the output gap.

### 4.3 Modelling Framework

The chapter explores inflation dynamics through a novel modelling framework that accommodates structural breaks by adopting model averaging techniques in a recursive set-up. The method involves estimation of a set of hybrid NKPCs, estimated over different sample sizes, providing the most reliable estimates of model parameters whatever sample window is used, accounting for the possibility that the responsiveness of inflation to expectations, past inflation and the real activity measure changes over time. In what follows we describe our modelling techniques and the construction of model weights.

Breaks due to policy change or changes in the appropriate measure of the real activity can be thought of as regime-type uncertainty and specification-type uncertainty. Both these types of uncertainty can be accommodated in a "meta-Phillips Curve," following the approach of Lee, Morley and Shields (2015). Specifically, we consider a set of hybrid NKPC models,  $M_{ijt}$ , each distinguished according to the measure of real activity,  $i$ , and the sample period for which the model is relevant. Specifically, the set of models characterising inflation dynamics over the period  $T_1 + j_{\max} - 1, \dots, T_n$  is given by:

$$\begin{aligned}
 M_{ijt} &: \pi_t = \lambda_{ijt} mc_{it} + \gamma_{fijt} E_t\{\pi_{t+1}\} + \gamma_{bijt} \pi_{t-1} + \varepsilon_{ijt} \\
 &: \text{where } mc_{it} \text{ is the real marginal cost measure} \\
 &: i = \begin{cases} 1 & \text{if the labour share is used} \\ 2 & \text{if the output gap is used} \end{cases} \\
 &: j = j_{\min}, \dots, j_{\max}, \quad t = T_1 + j_{\max} - 1, \dots, T_n,
 \end{aligned} \tag{4.5}$$

In any model, the subscript,  $i$ , indicates which measure of real activity used. The models are also distinguished by the time span over which a given Phillips Curve is assumed to hold, considered here to be in operation for  $j$  periods ending in period  $t$ . When there is a regime break the new period starts afresh so that the new model starts again. In practice, we might use a minimum sample size of 16 observations ( $j_{\min} = 16$ ), noting that the choice of the minimum regime length is driven by the need to have enough observations for estimation purposes. The maximum period for the survival of an unchanged inflation behaviour is theoretically unlimited. In practice, though, we expect inflation dynamics to change whenever there is a change in monetary policy stance which

is itself affected by the different Chairmen of the Federal Reserve. In the U.S. there have been six Federal Reserve Chairs since the mid-sixties so that, even in the absence of any other information, one might anticipate that there would be breaks every six or seven years and that a given policy rule would not last longer than ten years, i.e. using quarterly data  $j_{\max} = 40$ . Hence, for each measure of the real activity, there are  $40 - 16 + 1 = 25$  models that explain data ending at time  $T_1 + j_{\max} - 1$ , so that there are 50 candidate Phillips Curve models that differ according to the forcing variable and their relevant sample size. The next set of 50 models can be estimated for the period ending  $T_1 + j_{\max}$  and further sets of 50 models are estimated as we roll through the sample from  $T_1 + j_{\max} - 1$  to  $T_n$ , allowing for considerable flexibility in characterising regime change and shifts in the way the driver of inflation,  $i$ , is best measured. The left panel of Figure 4.1 shows the first set of candidate hybrid NKPC models which explain inflation dynamics at time  $T_1 + j_{\max} - 1$ . As we roll through the sample, we obtain the second set of candidate hybrid NKPC models that explain inflation at time  $T_1 + j_{\max}$  as shown in the right panel of figure 4.1. The estimated parameters in the individual hybrid NKPC,  $M_{ijt}$ , are denoted by  $\hat{\gamma}_{fijt}$ ,  $\hat{\gamma}_{bijt}$  and  $\hat{\lambda}_{ijt}$ .

The models  $M_{ijt}$  can be brought together in a "meta" model using methods based on Bayesian Model Averaging (BMA) techniques, allowing researchers to conduct econometric analysis without conditioning on a single model. The BMA attaches probabilities to models, or some parameters of interest, given the evidence found in the data. Seminal contributions by Draper (1995) and Raftery, Madigan and Hoeting (1997) examine the benefits of BMA extensively.

### 4.3.1 Bayesian Model Averaging

The considerable structural uncertainty surrounding inflation dynamics is reflected by the idea that inflation observed at time  $t$  could be explained by any of the  $2 \times (j_{\max} - j_{\min} + 1) = 2 \times 25 = 50$  different models according to (4.5) if we set  $j_{\min} = 16$  and  $j_{\max} = 40$  and use two measures of the real activity. Our meta-Phillips Curve accommodates regime uncertainty by using a weighted average of the model parameters in (4.5). Denoting the vector of parameters in the hybrid Phillips

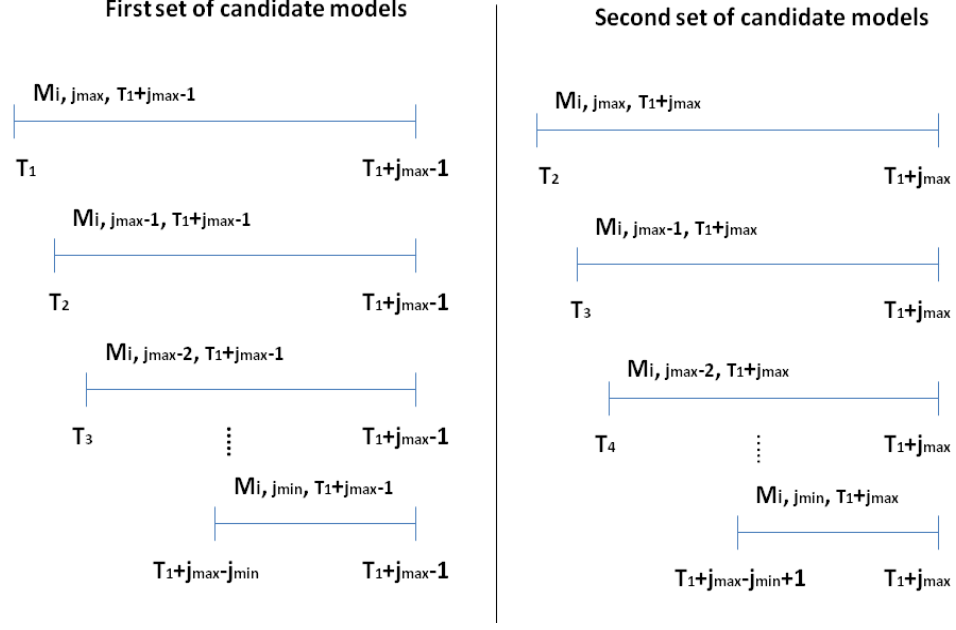


Figure 4-1: The first and second set of individual hybrid NKPC needed for the estimation of the meta-Phillips Curve.

Curve (equation (4.5)) at time  $t$  as:

$$\zeta_{ijt} = \begin{pmatrix} \lambda_{ijt} \\ \gamma_{fijt} \\ \gamma_{bijt} \end{pmatrix} \quad (4.6)$$

and noting that  $\mathbf{Z}_t = (\mathbf{z}_1, \dots, \mathbf{z}_t)$  represents all available information up to time  $t$ , our aim is to compute the average of the posterior probability of the parameters of interest i.e.  $\zeta_t$  under each model weighted by the corresponding posterior model probabilities. Using the BMA formula, taken from Draper (1995) and Hoeting et. al. (1999), this is given by:

$$\Pr(\zeta_{ijt}|\mathbf{Z}_t) = \sum_{i=1}^2 \sum_{j=16}^{40} \Pr(\zeta_{ijt}|M_{ijt}, \mathbf{Z}_t) \times \Pr(M_{ijt}|\mathbf{Z}_t), \quad (4.7)$$

Essentially,  $\Pr(\zeta_{ijt}|\mathbf{Z}_t)$  is the weighted average of the distributions of parameters in the hybrid new-Keynesian Phillips Curve given the individual models.  $\Pr(\zeta_{ijt}|M_{ijt}, \mathbf{Z}_t)$  is the distribution of  $\zeta_{ijt}$  on a specific model alone while the weights  $\Pr(M_{ijt}|\mathbf{Z}_t)$  is the posterior probability of model  $M_{ijt}$  given the data in the sample period reflecting how well model  $M_{ijt}$  fits the data. Since posterior model

probabilities add up to one  $\sum_{i=1}^2 \sum_{j=16}^{40} \Pr(M_{ijt}|\mathbf{Z}_t) = 1$ , they are treated as weights. The BMA formula deals with structural uncertainty embedded within  $\Pr(\zeta_{ijt}|\mathbf{Z}_t)$  by decomposing it into a weighted average of the conditional distributions (i.e. conditional on a specific model),  $\Pr(\zeta_{ijt}|M_{ijt}, \mathbf{Z}_t)$ , using as weights the posterior model probabilities,  $\Pr(M_{ijt}|\mathbf{Z}_t)$ .

### Conditional Distributions

A typical Phillips Curve analysis considers the first element on the right-hand side of (4.7) only, i.e.  $\Pr(\zeta_{ijt}|M_{ijt}, \mathbf{Z}_t)$ , working only with a particular model,  $(M^*)$ , which is assumed to be true and making inferences that are based on stochastic and parameter uncertainties. The conditional distribution can be approximated using the maximum likelihood (ML) estimator of the parameters in  $M^*$  and its associated density so that

$$\Pr(\zeta_{ijt}|M^*, \mathbf{Z}_t) = \int \Pr(\zeta_{ijt}|M^*, \zeta, \mathbf{Z}_t) \Pr(\zeta|M^*, \mathbf{Z}_t) d\zeta.$$

In the case of a standard linear regression model, we know that the departure of the ML estimated parameters,  $\hat{\zeta}_t^*$ , from their true value i.e.  $(\hat{\zeta}_t^* - \zeta_t|M^*, Z_t)$  follows a normal distribution  $N(0, \hat{V}_t^*)$  with mean zero and variance  $\hat{V}_t^*$  where  $\hat{V}_t^*$  is the ML estimated variance. Although  $\zeta_{ijt}$  is taken as fixed at the estimation stage, it can be viewed as a random variable at the inference stage, so that  $\Pr(\zeta_{ijt}|M^*, \mathbf{Z}_t)$  is approximated by  $N(\hat{\zeta}_{ijt}^*, \hat{V}_t^*)$  and standard inference carried out. Moreover, this simplification can be made for any model so that we can look at all 50 of our models of interest and base  $\Pr(\zeta_{ijt}|M^*, \mathbf{Z}_t)$  on the models' maximum likelihood estimates.

### The Model Weights

Model weights are constructed according to:

$$\begin{aligned} \Pr(M_{ijt}|\mathbf{Z}_t) &= \Pr(M_{ijt}|\mathbf{Z}_{t-1}, \mathbf{z}_t) \\ &\propto \Pr(\mathbf{z}_t|M_{ijt}, \mathbf{Z}_{t-1}) * \Pr(M_{ijt}|\mathbf{Z}_{t-1}) \\ &= \Pr(\mathbf{z}_t|M_{ijt}, \mathbf{Z}_{t-1}) * \sum_{k=1}^2 \sum_{l=16}^{40} \Pr(M_{ijt}, M_{klt-1}, \mathbf{Z}_{t-1}) * \Pr(M_{klt-1}|\mathbf{Z}_{t-1}). \end{aligned} \tag{4.8}$$

In practice, we can choose model weights so that they evolve over time, recursively updating them to reflect the extent to which they remain useful. A model's weight,  $\Pr(M_{ijt}|\mathbf{Z}_t)$ , depends on:



- the probability of observing the final observation,  $z_t$ , in the sample  $t-j, \dots, t$ , i.e.  $\Pr(\mathbf{z}_t|M_{ijt}, \mathbf{Z}_{t-1})$ , which, under standard normality assumptions, is proportional to the value of squared residuals at the end of the sample and
- the likelihood that the model remains relevant based on data up to  $t-1$ ,  $\Pr(M_{ijt}|\mathbf{Z}_{t-1})$ . This in turn depends on:
  - last period's weights,  $\Pr(M_{klt-1}|\mathbf{Z}_{t-1})$ , and
  - the transition probability,  $\Pr(M_{ijt}, M_{klt-1}, \mathbf{Z}_{t-1})$ . A simple structure for the transition probability is available if we assume that irrespective of inflation dynamics so far, there is a constant probability of a break in the way inflation behaves, probability  $\rho$ . If there is a break, inflation is assumed to enter a new regime starting again with the minimum sample size of 16 observations, using either measure of real activity with equal probability. That is,

$$\Pr(M_{ijt}|M_{klt-1}, \mathbf{Z}_{t-1}) = \begin{cases} 1 - \rho & \text{if there is no break in the PC} \\ \rho/2 & \text{if a break in the PC occurs.} \end{cases} \quad (4.9)$$

If inflation is explained by a previously estimated Phillips Curve, i.e. there is no break, the model just gets bigger by one additional observation while updating the weights on the different models recursively from one period to the next to reflect the likelihood that the models remain relevant. Thus, the transition probability is equal to  $1 - \rho$ . If a new Phillips Curve now explains inflation dynamics, such that a new inflation regime is "born", then the transition probability is equal to  $\rho/2$ . Notice that we divide by two because there are equal chances that the new Phillips Curve uses either measure of the real activity i.e. labour share and output gap. Taken together, (4.8) and (4.9) dictate the models' weights in each period. The models' weights for the first set i.e. the first period are assumed to be equal across all models.

This pragmatic approach to choosing weights and averaging over models accommodates both specification and structural uncertainty by combining individual Phillips Curves. The approach can capture the effect of complicated structural changes that are hard to disentangle using conventional

one-off structural breaks methods. The fact that model weights evolve over time allows for considerable flexibility in the way changes develop. In particular, the approach accounts for periods in which the responsiveness of inflation to the different factors changes both gradually from one state to another and abruptly. As Lee et. al. (2015) emphasise, the approach is more flexible than a standard time-varying parameter (TVP) model which deals with a pre-specified form of instability and more complex forms of instability require computationally more demanding estimation methods. The key advantages of his meta-modelling approach are, therefore, its easy implementation and provision of results that are easy to interpret and analyse.

The meta-Phillips Curve then consists of the individual estimated models, distinguished by the estimation period, sample size, and measure of real activity, and their weights and it is denoted by

$$\bar{M}_{T_n} = \{M_{ijt}, w_{ijt} \quad \text{for } i = 1, 2; j = 16, \dots, 40; t = 41, \dots, T_n\} \quad (4.10)$$

where  $w_{ijt}$  denotes the weight for model  $M_{ijt}$ , i.e.  $\Pr(M_{ijt}|\mathbf{Z}_t)$ . Since the estimated parameters in the individual hybrid NKPCs are given by  $\hat{\gamma}_{fijt}$ ,  $\hat{\gamma}_{bijt}$  and  $\hat{\lambda}_{ijt}$ , the estimated parameter on the forward-looking term in the meta-Phillips Curve is given by  $\bar{\gamma}_{ft} = \sum_{i=1}^2 \sum_{j=16}^{40} w_{ijt} \times \hat{\gamma}_{fijt}$ ; the estimated parameter on the backward-looking term is given by  $\bar{\gamma}_{bt} = \sum_{i=1}^2 \sum_{j=16}^{40} w_{ijt} \times \hat{\gamma}_{bijt}$ ; the estimated parameter on the forcing variable is given by  $\bar{\lambda}_t = \sum_{i=1}^2 \sum_{j=16}^{40} w_{ijt} \times \hat{\lambda}_{ijt}$  while the average sample size over a given Phillips Curve holds is given by  $\bar{j}_t = \sum_{i=1}^2 \sum_{j=16}^{40} w_{ijt} \times j_{ijt}$ .

### 4.3.2 Estimation

Since under rational expectations the error in the forecast of  $\pi_{t+1}$  is uncorrelated with information dated  $t$  and earlier, it follows from (4.4) that

$$E_t\{(\pi_t - \lambda mc_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1})\mathbf{z}_t\} = 0, \quad (4.11)$$

where  $\mathbf{z}_t$  is a vector of variables dated  $t$  and earlier so that it is orthogonal to the inflation shock in price  $t + 1$ . We use Generalised Methods of Moments (GMM) to estimate the models based on the orthogonality condition as given by equation (4.11). Note, that we also account for endogeneity in the forcing variable, since as noted by Roberts (1995), the marginal cost variable may be correlated with the error term since they could both be driven by cost-push shocks. In addition, measurement

errors in the construction of the unobservable output gap and more general the resource utilisation variable, call for using GMM.

Our estimation method aims to deal with issues raised in section 4.2.2. We particularly address the "weak instruments problem" and use a small instrument set, following the practice of Gali, Gertler and Lopez-Salido (2005). We use HAC, Newy-West type standard errors and with bandwidth (lag truncation parameter) equal to 2. (Usually the bandwidth is set equal to  $T^{\frac{1}{4}}$  where  $T$  is the sample size. In our framework, the sample size is variable, with maximum sample size of 40 observations, the rule requires a bandwidth set to 2.5 while with the minimum sample size of 16 observations, the rule requires a bandwidth equal to 2. Setting the bandwidth equal to 2, therefore, is a sensible choice). As a matter of fact, the approach is not really sensitive to the choice of bandwidth.

The construction of confidence bands for the meta-Phillips Curve is based on simulation methods in which artificial histories of the inflation series are generated to obtain the distribution of the meta-Phillips Curve estimated parameters. Full details are provided in appendix A.

## 4.4 Empirical Analysis

### 4.4.1 Data

We now turn to our empirical analysis of the U.S. inflation dynamics. Our dataset consists of U.S. aggregate time series at a quarterly frequency extending from 1950 : Q1 to 2012 : Q4 focusing on the variables used in Galí, Gertler and Lopez-Salido (2005) for comparability with their work. Series have been downloaded from the St. Louis Fed's FRED database and the U.S. Bureau of Labour Statistics. The data includes: the economy-wide GDP deflator; Non-farm business sector total compensation needed to compute labour share; hourly compensation and real output for the non-farm business sector; and commodity PPI (needed to construct commodity inflation). All growth rates are logarithmic and quarterly. For the construction of the output gap we use the log of real non-farm business sector output which is detrended using the HP filter with smoothing parameter equal to 1,600. We use the log of (demeaned) labour income share in the non-farm business sector and HP detrended log of real output in the non-farm business sector for measures of real activity.

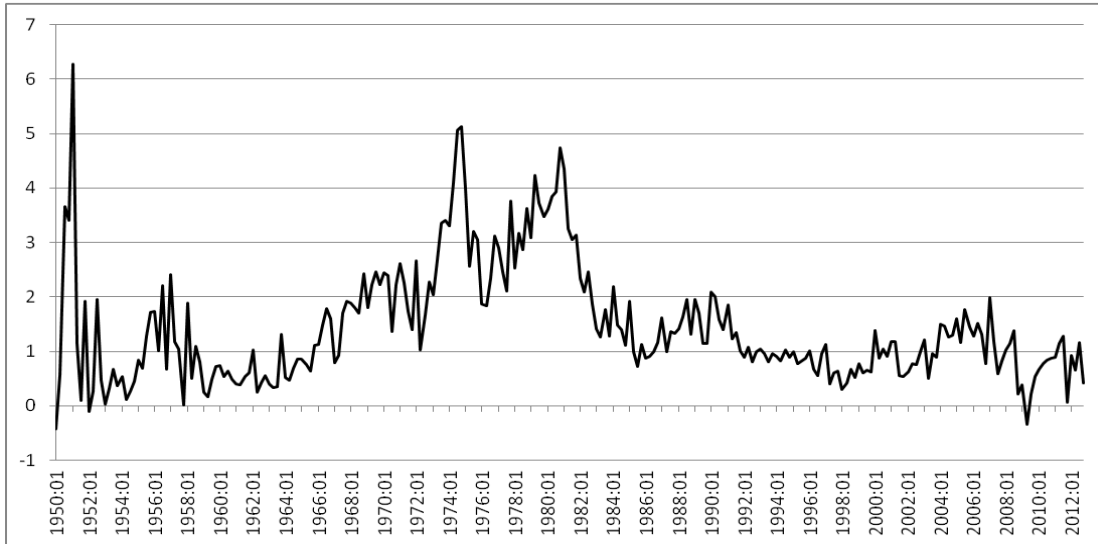


Figure 4-2: The U.S. Inflation Rate (Annualised (log) change of economy-wide GDP deflator).

Our measure of inflation is the economy-wide percent change in the GDP deflator. We use the overall deflator rather than the non-farm deflator for our analysis because we are interested in evaluating how well our model accounts for the movement in a standard broad measure of inflation. This series is plotted in Figure 4.2, demonstrating the dramatic changes in the inflation rate since the 1950s. Even from a mere observation, the researcher can identify the high inflation episodes and the switching inflation regimes. Accordingly, accounting for the potential structural instability in inflation dynamics is considered a key aspect of our empirical analysis.

#### 4.4.2 Results

Our characterisation of U.S. inflation is based on our estimated meta-Phillips Curve, obtained as a weighted average of the various models described in (4.5)<sup>6</sup>, obtained using the U.S. data for the period 1950 :  $Q1$ –2012 :  $Q4$ . The set of models considered is that described by (4.5), (4.8) and (4.9) and model weights are constructed accordingly. The instrument set used to estimate the models is the same as in Galí, Gertler and Lopez-Salido (2005) and includes four lags of inflation, two

---

<sup>6</sup>In order to account for non-zero steady-state inflation rate, we include an intercept in our models, following a number of studies including Russell et al. (2010) and Kim and Kim (2008). Russell et al. (2010) show that by including the intercept mean we accommodate structural breaks that exist in the mean of inflation.

lags of labour share, the output gap and wage inflation. To ensure sufficient degrees of freedom in estimating our Phillips Curves, we assume that regimes last a minimum of 4 years (i.e.  $j_{\min} = 16$ ) and regimes do not last longer than 10 years (i.e.  $j_{\max} = 40$ ) while we make use of two measures of the real activity. Given our setup, the first set of 50 Phillips Curves (25 models estimated over different sample sizes for each of the two forcing variables) that were estimated relate to the sample window of 40 observation from 1950 : Q1 – 1959 : Q4, estimating two curves (one using labour share and one using the HP filtered output gap) over the whole period, then two curves over the period 1950 : Q2 – 1959 : Q4, and so on, finishing with two models estimated over the minimum sample size of sixteen observations, i.e. over 1956 : Q1 – 1959 : Q4. Weights were calculated for each of these 50 models according to (4.8) and (4.9) which, among other things, also account for the ability of models to explain the final observation in 1959 : Q4. The second set of 50 Phillips Curves was estimated relating to the 40 observations from 1950 : Q2 – 1960 : Q1 and so on, moving recursively through the dataset.

The weighted average sample size is plotted in figure 4.3 and gives an idea of the timing of any inflation regime changes. The estimated meta-Phillips Curve parameters are plotted in figures 4.4-4.6 while figure 4.7 shows the cumulative sum of weights allocated to models that use the labour share as the forcing variable.

### **The eight inflation regimes**

Figure 4.2 shows the U.S. inflation rate, calculated based on the GDP deflator. The figure conveys the significant structural breaks the inflation process exhibited and demonstrates that there are identifiable periods of high and highly volatile inflation in which the behaviour of price-setting is likely to have changed. Continuity in the way inflation dynamics are characterised is reflected by a rising average sample size in figure 4.3. The bigger the average sample size, the longer the duration of a given hybrid NKPC. On the contrary, a sharp decline in the average sample size is a signal that inflation dynamics changed at that time. Accordingly, we can broadly identify eight inflation regimes, the start of which is captured by the sharp decrease in the average sample size. Table 4.1 summarises the key characteristics of these eight inflation regimes which are mostly driven by shifts in the conduct of monetary policy and significant developments in the economic environment.

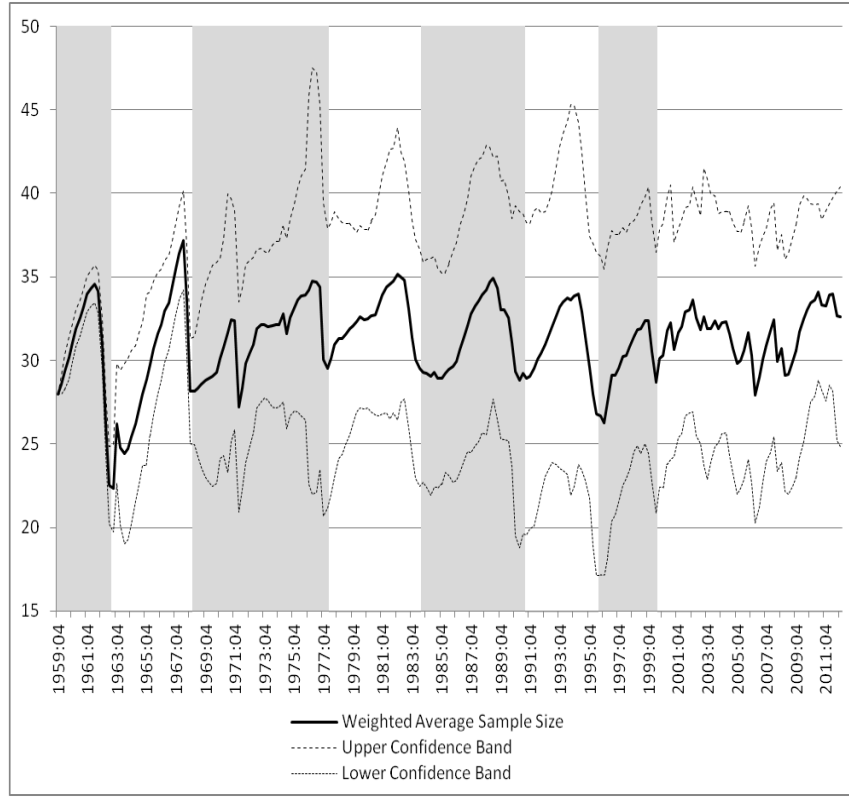


Figure 4-3: The evolution of the weighted average sample size,  $\bar{j}_t$ , over which a given hybrid NKPC holds. (Confidence bands constructed by bootstrap as plus/minus two times the coefficient's standard deviation).

In what follows we examine the evolution of the weighted average coefficients on the forward- and backward-looking terms and the forcing variable in the meta-Phillips Curve. As one might observe, the evolution of the average coefficient on the forward-looking term a mirror image of the evolution of the average coefficient on the backward-looking term. This is a by-product of the fact the we have restricted the sum of the two coefficients to sum to unity, which is a common assumption (see for example Galí and Gertler (1999)) that ensures that the discount factor is very close to one and that the long-run Phillips Curve is vertical. Here we focus attention on the behaviour of the coefficient on the forward-looking term since the two graphs tell the same story.

Table 4.1: The phases the meta-Phillips Curve has undergone through and their key characteristics		
Phase	Duration	Summary Overview
Martin/ Bretton Woods I	1959 : Q4–	Bretton Woods system of fixed exchange rates
	1963 : Q2	Strong tightening policy when inflationary pressures in place
		Mean Inflation = 0.53%. Variance of Inflation = 0.04.
Martin/ Bretton Woods II	1963 : Q3–	Rapid fiscal expansion
	1968 : Q4	Perceptual shift in policy making
		Fed pursuing goals other than dollar and price stability
		De-anchoring of inflation expectations
		Mean Inflation = 1.21%. Variance of Inflation = 0.33.
Burns and Miller/ The Great Inflation	1969 : Q1–	Collapse of Bretton Woods system
	1978 : Q1	High inflation episodes due to uncontrolled budget deficits and mounting oil prices
		Policy emphasis on stabilising real economic activity
		Mean Inflation = 2.06%. Variance of Inflation = 0.26.
Volcker I/ The Big Disinflation	1978 : Q2–	Major shift in U.S. monetary policy
	1984 : Q2	Proactive stance towards controlling inflation
		Mean Inflation = 2.89%. Variance of Inflation = 1.05.
Volcker II and Early Greenspan I/ The Onset of Great Moderation	1984 : Q3–	Strong anti-inflation credibility
	1991 : Q2	Well anchored expectations
		Mean Inflation = 1.39%. Variance of Inflation = 0.15.

Table 4.1 (Continued): The phases the meta-Phillips Curve has undergone through and their key characteristics		
Phase	Duration	Summary Overview
Early Greenspan II/ The deep recession	1991 : Q3– 1996 : Q2	Early 1990s prolonged deep recession
		Policy directed towards managing demand
		Inflation tamed in the fear of overheating
		Mean Inflation = 0.93%. Variance of Inflation = 0.02.
Mid-Greenspan/ The Boom Years	1996 : Q3– 2000 : Q2	First positive output gap after early 1990s recession
		Boom years until 2000.
		Shift of policy to pre-empt inflation
		Strong responses to inflation in the Taylor Rule
		Low and stabilised inflation
		Mean Inflation = 0.69%. Variance of Inflation = 0.08.
Late Greenspan/ Bernake The dot.com bubble and The Great Recession	2000 : Q3– 2012 : Q4	Early 2000s: outbreak of dot.com bubble
		High unemployment, substantial business failures
		Confidence plagued
		2007 – 2008: start of Great Recession
		Mean Inflation = 0.95%. Variance of Inflation = 0.20.

Note: Average inflation rate calculated using the (log of) economy-wide GDP deflator



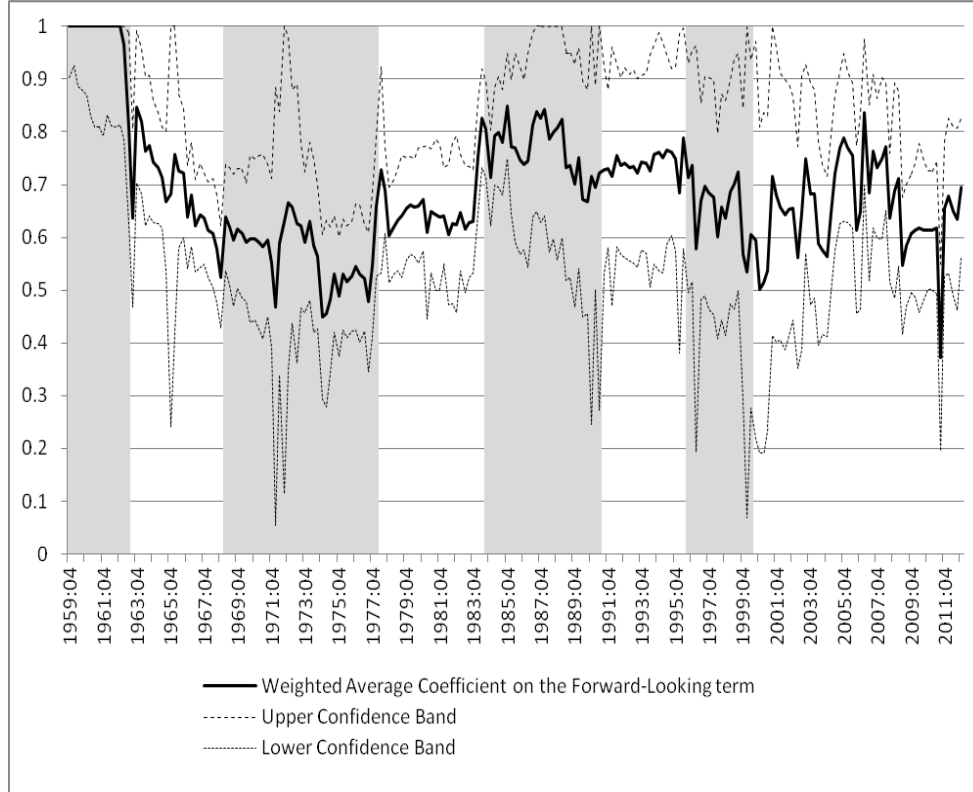


Figure 4-4: The evolution of the weighted average coefficient on the forward-looking term,  $\bar{\gamma}_{ft}$ , in the meta-Phillips Curve (4.10) as generated by recursive GMM estimation. (Confidence bands generated by bootstrap and constructed as plus/minus 2 times the coefficient's standard deviation).

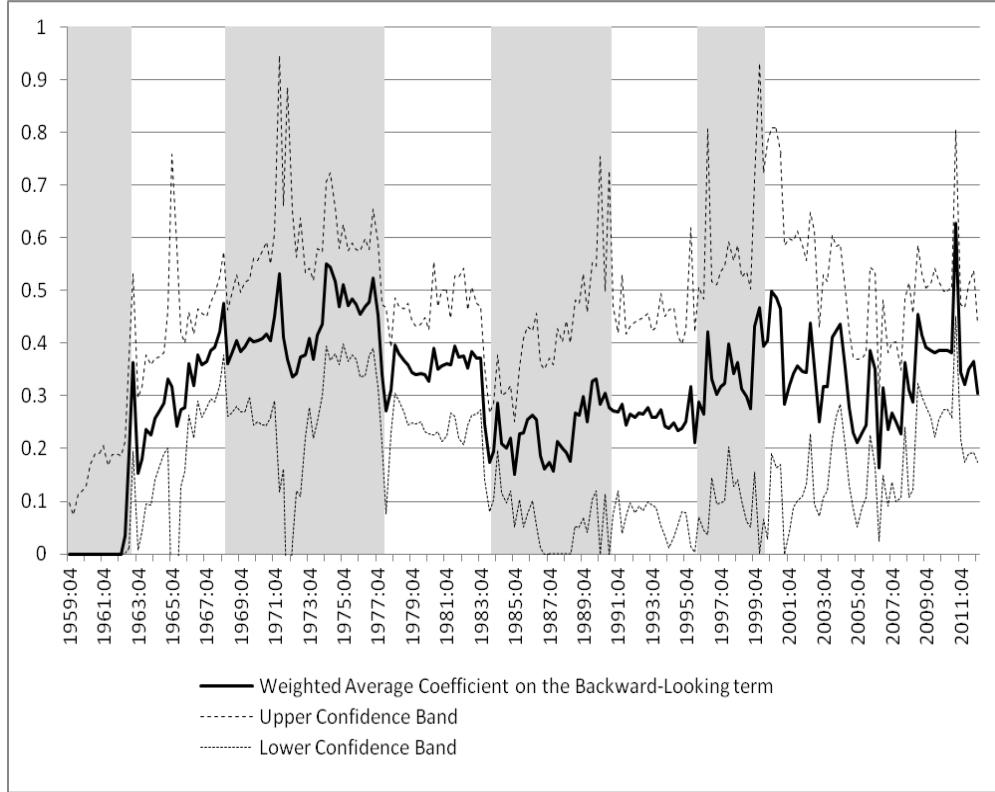


Figure 4-5: The evolution of the weighted average coefficient on the backward-looking term,  $\bar{\gamma}_{bt}$ , in the meta-Phillips Curve (4.10) as generated by recursive GMM estimation. (Confidence bands generated by bootstrap and constructed as plus/minus 2 times the coefficient's standard deviation).

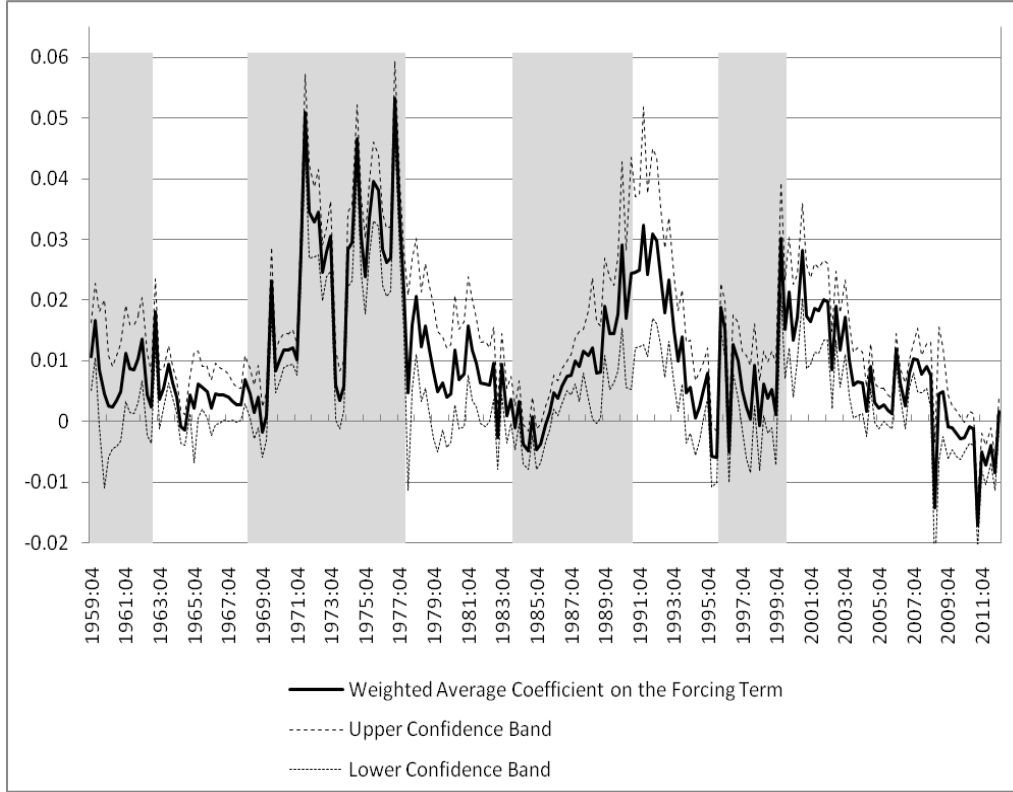


Figure 4-6: The evolution of the weighted average coefficient on the forcing variable,  $\bar{\lambda}_t$ , in the meta-Phillips Curve (4.10) as generated by recursive GMM estimation. (Confidence bands generated by bootstrap and constructed as plus/minus 2 times the coefficient's standard deviation).

Table 4.2: The average parameters in the meta-Phillips Curve in each phase				
Phase	Duration	Average $\bar{\gamma}_{ft}$	Average $\bar{\gamma}_{bt}$	Average $\bar{\lambda}_t$
Martin/ Bretton Woods I	1959 : Q4 – 1963 : Q2	0.985	0.0152	0.00751
Martin/ Bretton Woods II	1963 : Q3 – 1968 : Q4	0.688	0.3123	0.00468
Burns and Miller	1969 : Q1 – 1978 : Q1	0.569	0.431	0.0226
Volcker I	1978 : Q2 – 1984 : Q2	0.654	0.346	0.00848
Volcker II/ Early Greenspan I	1984 : Q3 – 1991 : Q2	0.766	0.234	0.00794
Early Greenspan II	1991 : Q3 – 1996 : Q2	0.741	0.259	0.0143
Mid-Greenspan	1996 : Q3 – 2000 : Q2	0.654	0.346	0.00814
Late Greenspan/ Bernake	2000 : Q3 – 2012 : Q4	0.652	0.347	0.00631

**Phase 1: Bretton Woods I** As table 4.1 summarises, 1960 – 1963 was a period of low and very stable inflation which is attributed to the stability established under the fixed exchange rate system. The monetary policy mechanism in place was automatic: Signs of overheated aggregate demand that threatened to accelerate inflation and undermine the country’s competitiveness were promptly addressed by triggering a strong tightening policy. The Fed’s commitment to maintain price stability reinforced its credibility and anchored inflation expectations. The fact that any shock had only transient effects meant that firms were more capable to predict the future prospects of inflation, so that the fraction of the backward-looking firms,  $\omega$ , was small. Since the parameters in the hybrid NKPC are functions of deep parameters, it is not surprising that the weighted average coefficient on the forward-looking term was very close to one, as shown in figure 4.4. Indeed, table 4.2, which shows the average estimates on the meta-Phillips Curve parameters over the different inflation regimes, suggests that in this first regime inflation dynamics could be described by a purely forward-looking Phillips Curve, as average  $\bar{\gamma}_{ft}$  was as high as 0.985. The weighted average coefficient on the forcing variable over this period was relatively stable and remained at very low levels, as shown in figure 4.6.

**Phase 2: Bretton Woods II** In the second inflation regime, inflation doubled and became much more volatile as conveyed by table 4.1. Bordo and Eichengreen (2008) emphasise that in 1963, there has been an important perceptual shift in the assumed responsibilities of the Fed that

considered itself free to pursue goals other than dollar stabilisation, undermining the importance of controlling inflation. Effectively, policymakers placed high importance on stabilising the real economic activity and paid much less attention to price stability, unmooring inflation expectations (Orphanides and Williams (2012)). The uncertain prospects for the inflation and the loss of credibility for the Central Bank meant that firms became less forward-looking when setting their prices, and the fraction of backward-looking firms,  $\omega$ , increased. This is reflected by the drop in coefficient on the forward-looking term in the meta-Phillips Curve. Table 4.2 shows that average  $\bar{\gamma}_{ft}$  dropped to 0.688, while the average  $\bar{\lambda}_t$  did not exhibit any major shifts. Despite the adverse developments, the forward-looking term remained dominant.

**Phase 3: The Great Inflation** The third regime, starting from 1969 :  $Q1$  and ending in 1978 :  $Q1$ , was marked by unusual economic turmoil. The collapse of the Bretton Woods system plagued inflation expectations while the oil price shocks brought inflation to unprecedentedly high levels. The high and highly volatile inflation environment did not leave price-setter unaffected. Ball, Mankiw and Romer (1988) show that low and stable inflation leads to less-frequent price adjustments (Ball-Mankiw-Romer conjecture). The idea is that low-inflation allows firms to leave their prices fixed at little cost. Accordingly, the Great Inflation forced firms into more frequent price adjustments, so that the proportion of firms that left their prices unchanged,  $\theta$ , dropped significantly. This change in the deep parameter meant that the coefficient on the forward-looking term in the hybrid NKPC exhibited a downward path, while the coefficient on the forcing variable increased substantially. Table 4.2 validates the argument since average  $\bar{\gamma}_{ft}$  dropped to its lowest levels during the Great Inflation period, while average  $\bar{\lambda}_t$  more than tripled. The steepening of the meta-Phillips Curve over this period is also evident in figure 4.6.

**Phase 4: The Big Disinflation** The period 1978 :  $Q1$  – 1984 :  $Q2$  remained in history as the Big Disinflation. By the end of this period, Volcker managed to bring inflation down substantially. In contrast to policies in the previous years, Volcker’s programme involved a proactive stance towards controlling inflation. The meta-approach has successfully identified this major shift since the weighted average coefficient on the forward-looking term escaped from the 0.5 neighbourhood and remained as high as 0.65 during Volcker years. The average coefficient on the forcing variable dropped significantly in response to policy changes. Our approach, therefore, validates a number

of studies that found a structural break around 1982/3 (Ball, Mankiw and Romer (1988), Zhang, Osborn and Kim (2008), and Fuhrer, Olivei and Tootell (2009)).

**Phase 5: The Onset of Great Moderation** The late 1980s period was considered to be the onset of the Great Moderation where output volatility reduced substantially after the government's economic stabilisation policy. This was the time when the Central Bank regained back its credibility after inflation was tamed, while its wider independence meant that monetary policy became free from fiscal concerns. Changes in the way Fed communicated monetary policy plans brought greater transparency resulting in more effective policies. Accordingly, confidence in the Central Bank and the future prospects of inflation allowed firms to become more forward-looking, decreasing the proportion of backward-looking firms,  $\omega$ , causing  $\bar{\gamma}_{ft}$  in the meta-Phillips Curve to rise. Indeed, table 4.2 shows that average  $\bar{\gamma}_{ft}$  over this period was as high as 0.766, and this increase becomes apparent in figure 4.4.

The decline in inflation persistence<sup>7</sup> in late 1980s, as documented by a number of studies, (Taylor (2000), Cogley and Sargent (2001) and Levin and Piger (2004)) can also explain the increase in the weighted average coefficient on the forward-looking term. Roberts (2006) explains that when inflation is persistent, agents use the lagged inflation as predictor of future inflation so that the proportion of backward-looking firms,  $\omega$ , is large. The decline in inflation persistence is thought to have decreased  $\omega$ , putting an upward pressure on  $\bar{\gamma}_{ft}$ .

Although table 4.2 indicates that on *average* coefficient  $\bar{\lambda}_t$  did not experience major shifts relative to previous regime, figure 4.6 conveys an important message: During this regime the slope of the meta-Phillips Curve has switched sign, becoming positive and steep again. This can be attributed to the decline in the proportion of backward-looking firms,  $\omega$ , which determines the level of  $\gamma_f$  in the hybrid NKPC, as shown in (4.4).

**Phase 6: Recessionary Years** Despite that this regime was marked by a prolonged and deep recession that brought about high unemployment, inflation was kept under control. Table 4.1 indicates that inflation was maintained at very stable and low levels. According to Ball-Mankiw-Romer conjecture low and stable inflation leads to less frequent price adjustments, so that the

---

<sup>7</sup>Inflation persistence is defined as the speed with which inflation returns to baseline after a shock.

fraction of firms that leaves prices unchanged,  $\theta$ , increases. This is reflected in the coefficients of the meta-Phillips Curve, and particularly the weighted average coefficient on the forcing term that dropped significantly as a result of the increase in  $\theta$ . Although the weighted coefficient on the forward-looking term,  $\bar{\gamma}_{ft}$ , did not exhibit too much variation relative to the previous period, it remained at high levels, echoing the infrequent price adjustments.

**Phase 7: The Boom Years** The mid-Greenspan phase was characterised by a shift of the policy to pre-empt inflation involving strong responses to inflation in the Taylor rule, while shifting attention towards responding more rigorously to output gap. This was the time when the output gap became positive for the first time after the early 1990s recession. The meta-approach has successfully captured this change in policy, as shown by the declining weighted average coefficient on the forward-looking term, as shown in figure 4.4. The weighted average coefficient on the forcing variable remained relatively stable on average but exhibited highly volatile behaviour over the interval.

**Phase 8: The dot.com Bubble and Great Recession** The final inflation regime was characterised by high uncertainty due to the dot.com bubble and a number of business failures in early 2000s which mounted in 2007 – 2008 with the outbreak of the global financial crisis and a series of bank failures. Unlike the 1970s crisis, inflation was not a worry since with interest rates hitting the zero lower bound policymakers were more concerned on how to avoid deflation. The high uncertainty, in combination with fears of uncontrolled budget deficits, de-anchored inflation expectations and brought unrest among price-setters, consumers and investors. The erratic and highly volatile path that the weighted average coefficient on the forward-looking term experienced reflects the uncertain times. On average, the average  $\bar{\gamma}_{ft}$  over the period remained high at 0.652, as shown in table 4.2.

The most important change during this periods is the well-documented flattening of the Phillips Curve as seen by the steady drop in the forcing variable coefficient,  $\bar{\lambda}_t$ . A mere observation of figure 4.6 reveals that there has been an important structural break in early 2000s with the weighted average coefficient on the forcing variable declining steadily and reaching negative values, although confidence bands included positive values. The Ball-Mankiw-Romer conjecture can partly explain the flattening of the hybrid NKPC: The low and stable inflation environment led to less frequent

price adjustments, and an increase in  $\theta$ , reducing the slope of the Phillips Curve. Blanchard and Galí (2007) attribute the flattening of the Phillips Curve to globalisation and the reduction in the pass-through of oil prices to prices charged to consumers. The decrease in firms' pricing power and global competition results in more firms leaving their prices unchanged ( $\theta$  increases).

### **The dominance of the forward-looking term**

Figures 4.4 and 4.5 demonstrate that the forward-looking component is more dominant than the backward-looking component for the determination of inflation dynamics. With few exceptions, the weighted average coefficient on the forward-looking term was above 0.5, suggesting that expectations are important drivers of current inflation and the characterisation of inflation dynamics based purely on a backward-looking component is surely misleading.

Russell et al. (2010) show that if structural breaks are not accounted for, estimated coefficients on the dynamic inflation terms will be biased. More importantly, the authors illustrate that once shifts in the mean rate of inflation have been addressed, expected inflation in the hybrid NKPC becomes insignificant, arguing that price-setting agents are not as forward-looking as modelled in the hybrid theories. While we show that, depending on the prevailing economic conditions and the monetary policy in place, price-setting agents may become more or less forward-looking, we find that expected inflation plays a dominant role in inflation dynamics, even when structural breaks are incorporated in estimation and analysis. This finding validates a number of studies that support the view that the forward-looking term is important (see for instance, Galí and Gertler (1999), Rotemberg and Woodford (1997)).

### **Labour share vs Output Gap**

Figure 4.7 shows the distribution of weights among models that use the labour share as the forcing variable and models that use the HP filtered output gap. With very few exceptions, labour share models attract the highest portion of weights, sometimes receiving almost the whole share of weights, (for example, 2001) with 99% of weights favouring labour share as the forcing variable. The graph validates Galí and Gertler's (1999) argument that output gap is not an appropriate measure of real activity and labour share appears to explain current inflation better than output gap.



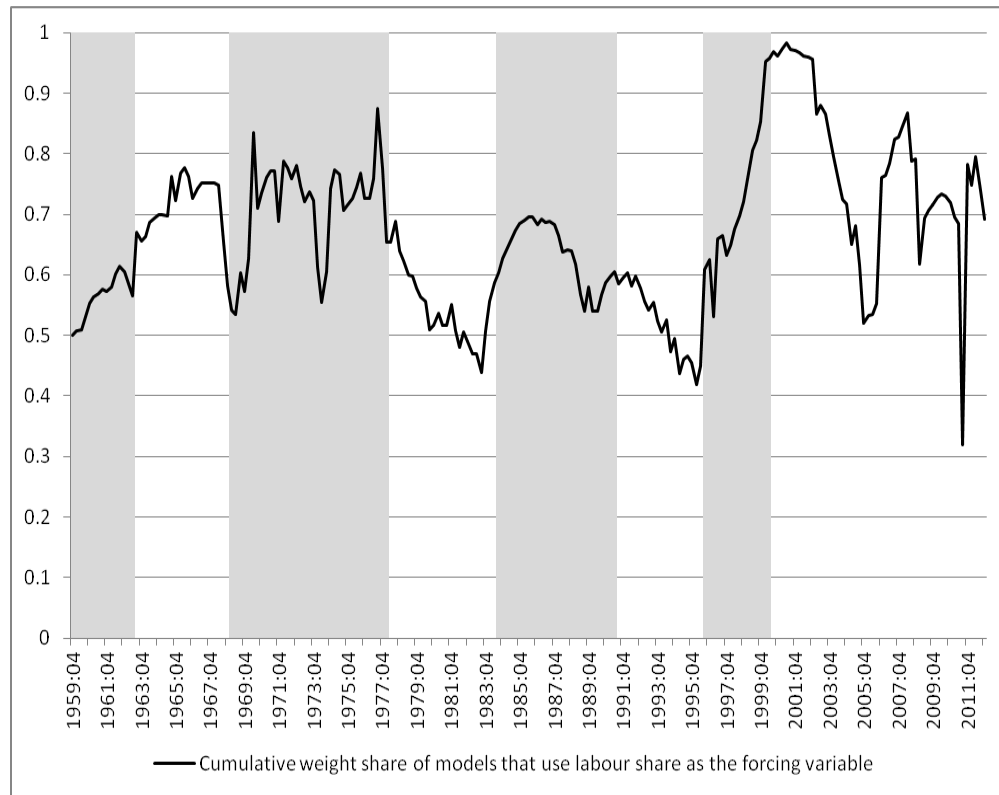


Figure 4-7: The evolution of weights allocated to models that use labour share as the forcing variable.

As the authors argue (1999, p.197), labour share “directly accounts for the impact of productivity gains on inflation, a factor that simple output gap measures often miss.”

One additional observation is that the share of models that use the labour share increases whenever the weighted average coefficient on the backward-looking term in the meta-Phillips Curve increases. One possible interpretation is that the labour share is a more immediate/ real-time measure of the marginal cost, while the HP output gap is more forward-looking, being based on both past and future values of the real GDP.

#### 4.4.3 Robustness Check

##### Using a larger and a smaller instrument set

In section 2.3.3 we have discussed the weak instruments and identification problems that usually arise in empirical Phillips Curve analysis. The idea that weak instruments lead to imprecise model

estimates is carefully incorporated in our chosen weight structure. Model weights depend on the likelihood of observation in the sample which is proportional to the value of squared residuals at the end of the sample for each model. Models that strongly influenced by the weak instruments problem are badly estimated so that they produce large squared residuals, resulting in smaller weights.

An important implication of the weak instruments problem is that estimates are very sensitive to the choice of the instrument set. Effectively, very different parameter estimates are usually obtained when a different instrument set is used in estimation. Our meta approach is shown to lessen the weak-instruments problem since the choice of different instrument set does not appear to alter the resulting estimates and inflation dynamics. Figures B4.1-B4.5 in Appendix B provide the evolution of the weighted coefficients of the meta-Phillips Curve, together with the weighted average sample size and cumulative sum of weights allocated to models that used the labour share as the forcing variable, based on an expanded instrument set that includes four lags of inflation, four lags of the labour share, two lags of the HP filtered output gap, two lags of wage inflation and two lags of PPI commodity inflation. The graphs look very similar to figures 4.3-4.7 based on the benchmark instrument set, illustrating that the approach is not sensitive to the choice of instrument set. Similarly, Figures C4.1-C4.5 in Appendix C, showing results based on a smaller instrument set, which includes two lags of inflation, two lags of labour share, two lags of the output gap and two lags of wage inflation, are very similar to graphs 4.3-4.7 based on the benchmark instrument set. Inflation regimes exactly coincide with those obtained in the benchmark case, and the interpretation of results hardly changes.

### **Choosing a different probability of break**

Figures D4.1-D4.5 suggest that the meta-Phillips Curve is not affected by the choice of the probability of a break. The benchmark case assumed that the probability of breaks is 0.01. In their paper, Russell et al. (2010) find that U.S. inflation exhibited nine breaks over a period spanning 50 years, implying a probability of breaks of 0.05<sup>8</sup>. Figures D4.1-D4.5 were based on the assumption

---

<sup>8</sup>9 breaks in 200 quarters, implies that the probability of break is  $9/200 \simeq 0.05$ .

that the probability of break is 0.05. While the figures showing the evolution of weighted average coefficients are very similar to the benchmark case, figure D4.1 which plots the weighted average sample size exhibits some variation. Despite the fact that the shape of D4.1 is almost identical to figure 4.4, the entire graph appears to have shifted down. While the location of breaks is not affected by the larger probability of break, the duration of any given hybrid NKPC is shown to decrease. This is intuitive: Higher probability of break implies that on average, the expected duration of a given Phillips Curve becomes shorter.

## 4.5 Conclusion

This chapter exploits model averaging in the context of behavioural modelling and inference in order to characterise inflation dynamics since the 1960s in a very flexible way. The model averaging technique is employed so that the modeller can overcome the regime uncertainty related to changing monetary policies and economic conditions. Our findings are threefold: First, we find that the combined Phillips Curve provides a flexible but compelling characterisation of inflation dynamics in the United States over the last fifty years with no single Phillips Curve dominating at any point in time. The combined Phillips Curve captures important shifts in the conduct of monetary policy and highlights key changes in inflation dynamics. The eight inflation regimes, identified by the meta-approach, reflect eminent developments ranging from the collapse of the Bretton Woods System, the oil-price shocks of the 1970s, the Great Moderation and the well-documented flattening of the Phillips Curve.

Second, our findings make important contributions to the ongoing debate about which of the forward-looking and backward-looking component dominates inflation dynamics. The meta-Phillips Curve accommodates breaks and so, if studies that demonstrate that the forward-looking component is a manifestation of unaccounted breaks (see for example Russell et al. (2010)) the coefficient on the forward-looking component should have been close to zero. We discover that even when accounting for structural breaks in the Phillips Curve relationship, the forward-looking term is still the dominant term, validating a number of studies that suggest that forward-looking expectations are important drivers of current inflation.

Third, we find that although the usefulness of labour share vis-a-vis the HP detrended output gap changes over time, on average, labour share appears to be a better measure of the marginal cost and the real activity since models that use the labour share as the forcing variable attract a higher proportion of weights across time. This validates a number of studies which demonstrate the superiority of labour share in allowing both productivity and wage pressures to influence inflation.

## 4.6 Appendix A: Confidence bands for estimated parameters in the meta-Phillips Curve

The construction of the confidence bands involves three steps:

(i) Generate  $S$  artificial series for inflation, denoted by  $\pi_t^{(s)}$   $s = 1, \dots, S$ ;  $t = T_1, \dots, T_n$  where  $S = 10,000$ . The simulated inflation series, obtained through the  $s^{th}$  replication is based on a particular NKPC,  $M_{ijt}$ , and is computed by:

$$\pi_t^{(s)} = \hat{\lambda}_{ijt}mc_{it} + \hat{\gamma}_{fijt}\tilde{E}_t\{\pi_{t+1}^{(s)}\} + \hat{\gamma}_{bijt}\pi_{t-1}^{(s)} + \varepsilon_{ijt}^{(s)} \quad (4.12)$$

where the error  $\varepsilon_{ijt}^{(s)}$  is computed parametrically based on a random draw from a standard normal distribution. In particular,  $\varepsilon_{ijt}^{(s)}$  are found by multiplying the random number drawn from the normal distribution by the standard error of the regression from the original estimated model,  $M_{ijt}$ . Since, for each replication, the inflation series is computed recursively, we cannot use the future values of the simulated inflation rates, i.e.  $E_t\{\pi_{t+1}^{(s)}\}$ , as they are unknown. Instead, we use the fitted values of  $\pi_{t+1}$ , from the first stage regression of actual future inflation on the instruments (denoted by  $\tilde{E}_t\{\pi_{t+1}^{(s)}\}$ ).

(ii) Estimate a meta-Phillips Curve model,  $\bar{M}_{T_n}^{(s)}$ , for each of the simulated inflation histories, thereby obtaining a set of estimates of the individual NKPCs parameters, denoted by  $\hat{\lambda}_{ijt}^{(s)}$ ,  $\hat{\gamma}_{fijt}^{(s)}$  and  $\hat{\gamma}_{bijt}^{(s)}$ , as well as the weighted average parameters that characterise the meta-Phillips Curve  $\bar{\lambda}_t^{(s)}$ ,  $\bar{\gamma}_{ft}^{(s)}$  and  $\bar{\gamma}_{bt}^{(s)}$ .

(iii) The set of the meta Phillips Curve parameters,  $\bar{\lambda}_t^{(s)}$ ,  $\bar{\gamma}_{ft}^{(s)}$  and  $\bar{\gamma}_{bt}^{(s)}$  across  $s = 1, \dots, S$  provides a distribution of estimated values. Confidence bands are then constructed as two times the standard errors of estimated parameters, each side of the mean estimates.

To generate each simulated history of  $\pi_t^{(s)}$   $s = 1, \dots, S$ ;  $t = T_1, \dots, T_n$  in step (i) we follow two stages. In stage 1,  $j_{max} - 1 = 39$  "initial" simulated observations ( $\pi_t^{(s)}$  for  $t = T_1, \dots, T_{39}$ ) are generated in the absence of any information on the probability of any one of the 50 candidate models being true. In stage 2, the remaining simulated observations ( $\pi_t^{(s)}$  for  $t = T_{40}, \dots, T_n$ ) take account of the probability that the data is generated by model  $M_{ijt}$  which is subject to structural breaks.

In detail, in stage 1, for the generation of *each* of the first  $j_{max} - 1 = 39$  observations i.e. for the time  $t = T_1, \dots, T_{39}$ , we use equation (4.12) based on a randomly chosen model out of the 50 models in the first set of candidate hybrid NKPCs.

In stage 2, the generation of the remaining observations reflects that at each point in time there is a constant probability of break. For the generation of the  $(j_{max})^{th} = 40^{th}$  observation of the inflation series for time  $T_1 + j_{max} - 1$ , it is assumed that inflation enters a new regime. Consequently, the 40<sup>th</sup> observation is generated according to (4.12) based on either the estimated model  $M_{1,16,T_1+j_{max}-1}$  or  $M_{2,16,T_1+j_{max}-1}$ . To determine whether we should use the labour share model or the output model, we take a random draw from a uniform distribution. If this is greater than 0.5 then the labour share model is used and the output gap model is used otherwise. For the generation of the 41<sup>st</sup> observation of the inflation series for time  $T_{41} = T_1 + j_{max}$ , we acknowledge the possibility that a break can take place. As a result two cases arise:

(a) If a break occurs, the 41<sup>st</sup> observation is based on either model  $M_{1,16,T_1+j_{max}}$  or  $M_{2,j_{min},T_1+j_{max}}$  (whether the labour share or output gap model is used is determined like before).

(b) If there is regime continuity, the 41<sup>st</sup> observation is generated according to either model  $M_{1,17,T_1+j_{max}}$  if model  $M_{1,16,T_1+j_{max}-1}$  was used to generated the previous observation or model  $M_{2,17,T_1+j_{max}}$  if model  $M_{2,16,T_1+j_{max}-1}$  was used to generated the previous observation.

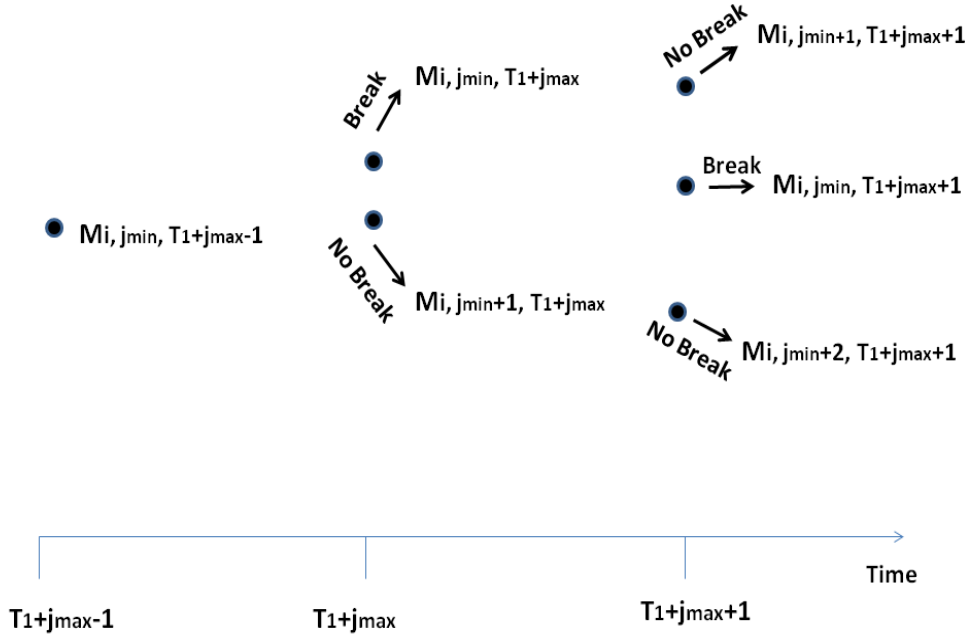


Figure A1: The construction of confidence bands, accounting for the constant probability of break at each point in time.

Figure A1 summarises the contingencies that occur when there is a constant probability of break at each point in time. To determine whether a break takes place, a random number from the uniform distribution is drawn. If this number is smaller than the probability of a break, then we assume that a break occurs and the new observation is generated according to (a) above. The procedure continues until we reach the end of the sample, where we generate the last observation for time  $T_n$ .

Once the first simulated history of the inflation series is obtained, we estimate the meta-Phillips Curve, following the methodology described in sections 4.3.1-4.3.3, and obtain the GMM estimates  $\bar{\lambda}_t^{(s)}$ ,  $\bar{\gamma}_{ft}^{(s)}$  and  $\bar{\gamma}_{bt}^{(s)}$  for  $s = 1$ . The procedure is repeated 10,000 times, obtaining for each  $s$ , the meta-Phillips Curve estimates  $\bar{\lambda}_t^{(s)}$ ,  $\bar{\gamma}_{ft}^{(s)}$  and  $\bar{\gamma}_{bt}^{(s)}$  thereby obtaining a distribution of the weighted coefficient estimates which allows us to determine the standard deviation of the estimated weighted average coefficients. Confidence bands are then constructed by multiplying the standard deviation of each weighted average coefficient by two.

## 4.7 Appendix B: Larger Instrument Set

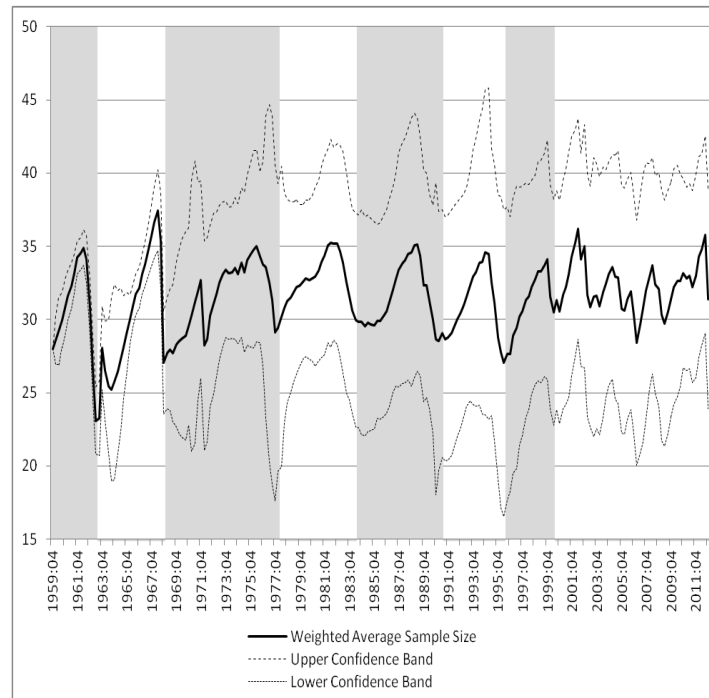


Figure B4.1: The evolution of the weighted average sample size,  $\bar{j}_t$ , over which a given hybrid NKPC holds, using a larger Instrument Set.

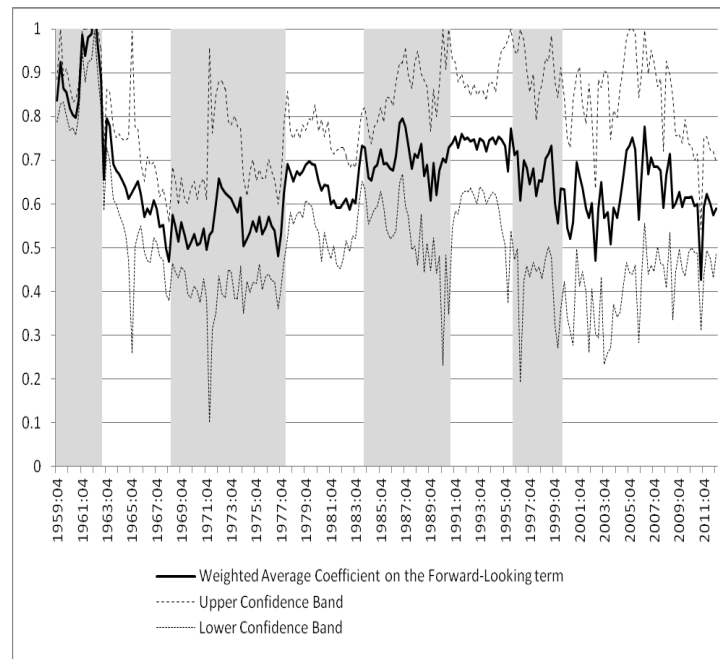


Figure B4.2: The evolution of the weighted average coefficient on the forward-looking term,  $\bar{\gamma}_{ft}$ , in the meta-Phillips Curve as generated by recursive GMM estimation, using a larger Instrument Set.

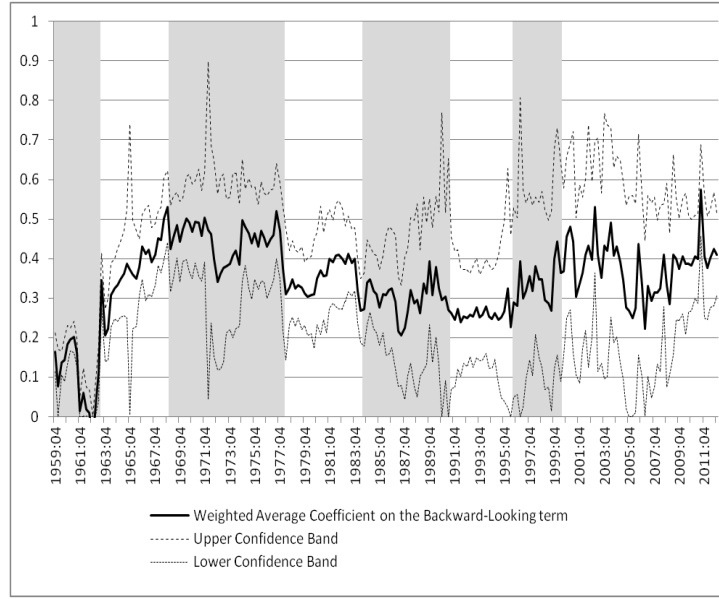


Figure B4.3: The evolution of the weighted average coefficient on the backward-looking term,  $\bar{\gamma}_{bt}$ , in the meta-Phillips Curve as generated by recursive GMM estimation, using a larger Instrument Set.

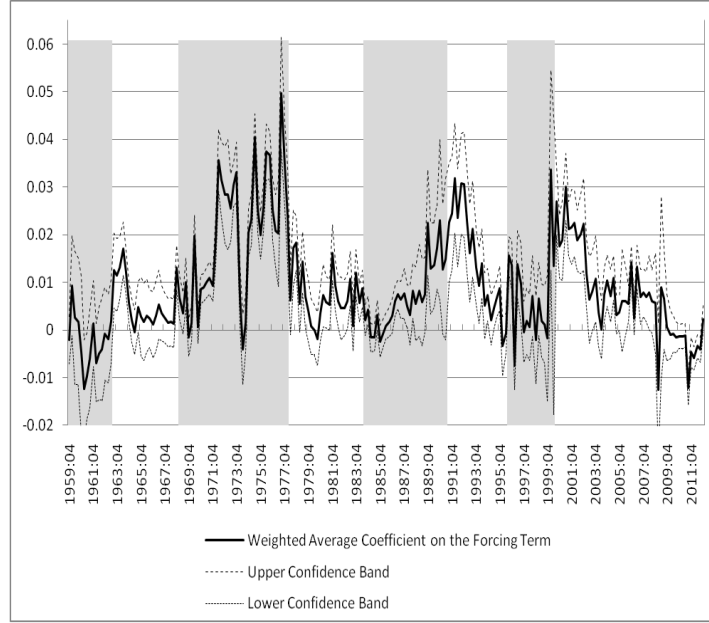


Figure B4.4: The evolution of the weighted average coefficient on the forcing variable,  $\bar{\lambda}_t$ , in the meta-Phillips Curve as generated by recursive GMM estimation, using a larger Instrument Set.



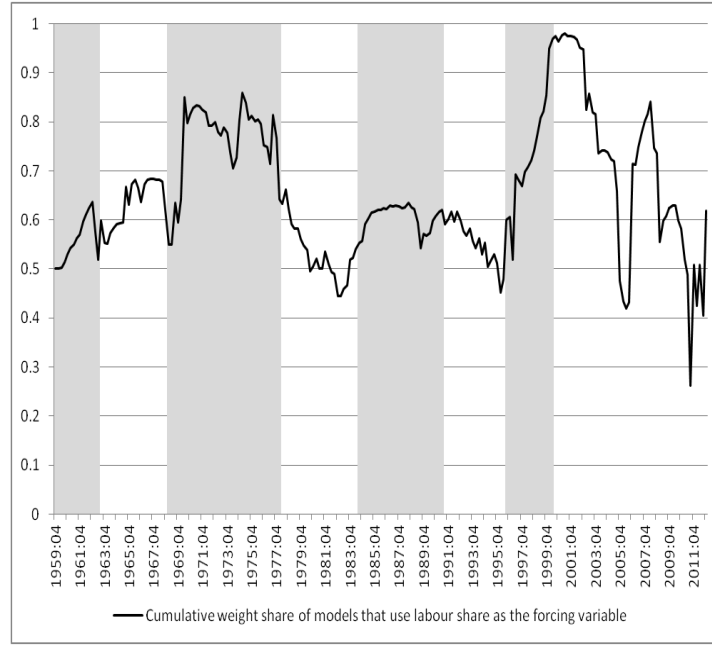


Figure B4.5: The evolution of weights allocated to models that use the labour share as the forcing variable, based on a larger Instrument Set.

## 4.8 Appendix C: Smaller Instrument Set

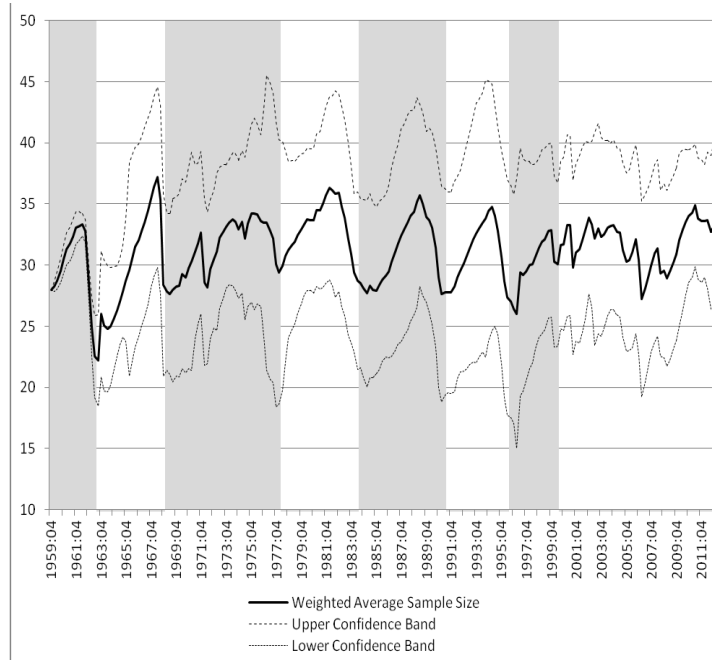


Figure C4.1: The evolution of the weighted average sample size,  $\bar{j}_t$ , over which a given hybrid NKPC holds, using a smaller Instrument Set.

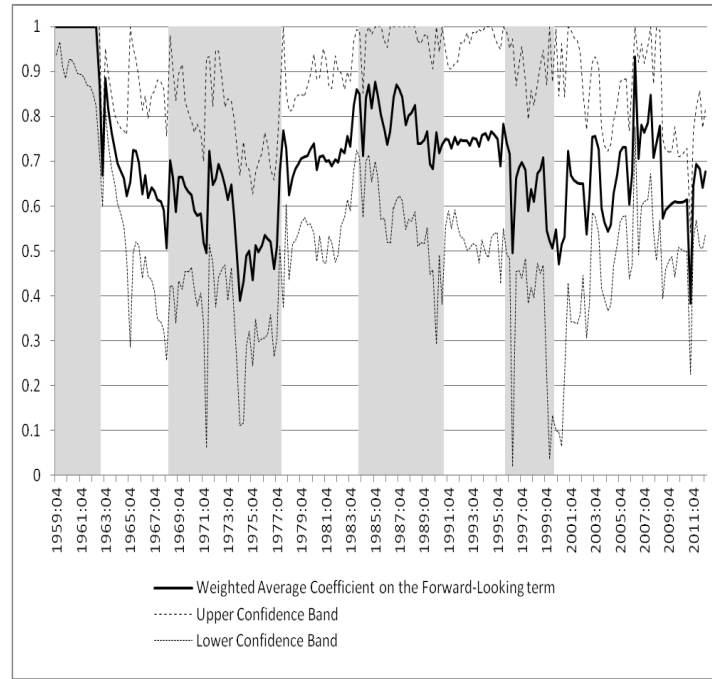


Figure C4.2: The evolution of the weighted average coefficient on the forward-looking term,  $\bar{\gamma}_{ft}$ , in the meta-Phillips Curve as generated by recursive GMM estimation, using a smaller Instrument Set.

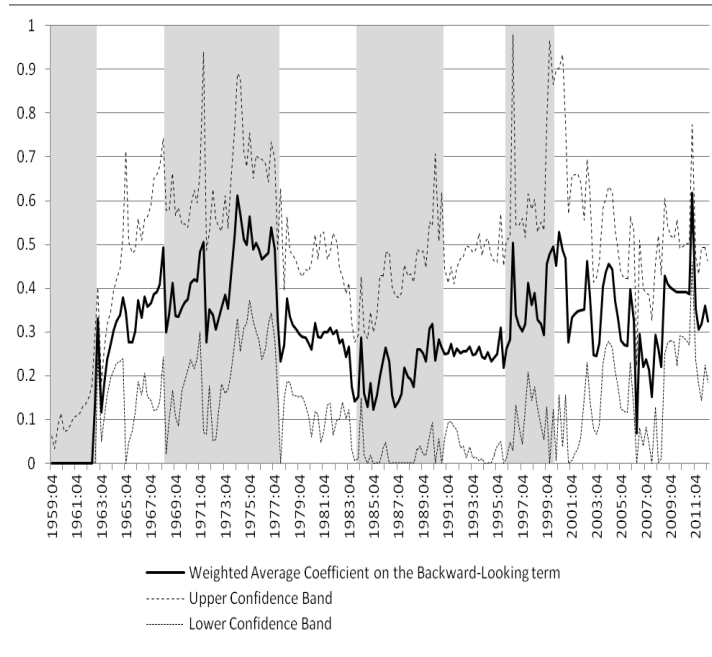


Figure C4.3: The evolution of the weighted average coefficient on the backward-looking term,  $\bar{\gamma}_{bt}$ , in the meta-Phillips Curve as generated by recursive GMM estimation, using a smaller Instrument Set.

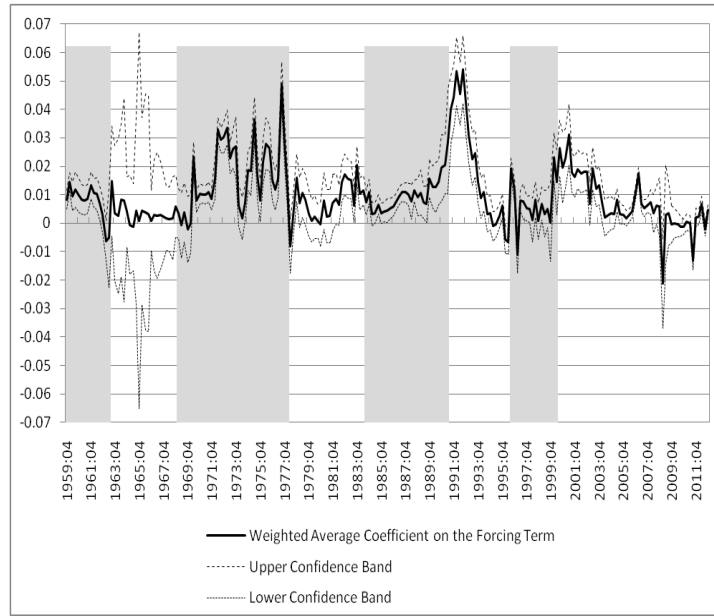


Figure C4.4: The evolution of the weighted average coefficient on the forcing variable,  $\bar{\lambda}_t$ , in the meta-Phillips Curve as generated by recursive GMM estimation, using a smaller Instrument Set.

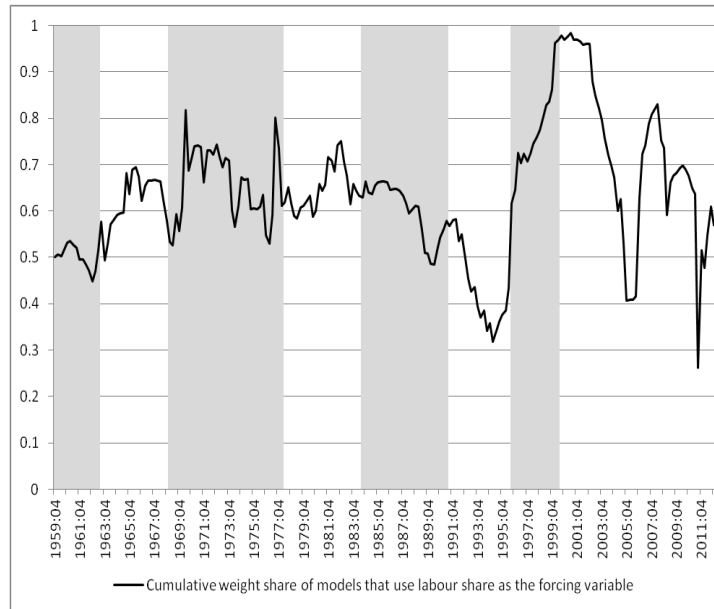


Figure C4.5: The evolution of weights allocated to models that use the labour share as the forcing variable, based on a smaller Instrument Set.

## 4.9 Appendix D: Larger Probability of Break

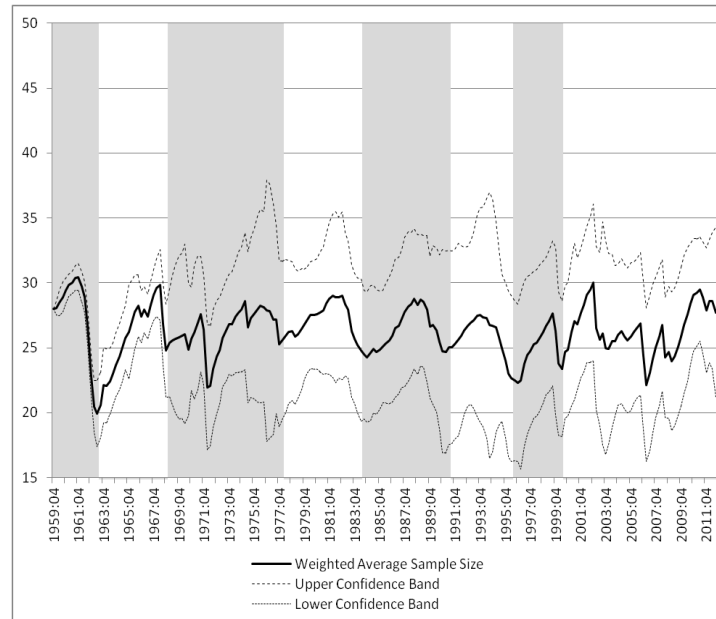


Figure D4.1: The evolution of the weighted average sample size,  $\bar{j}_t$ , over which a given hybrid NKPC holds, based on a higher probability of break.

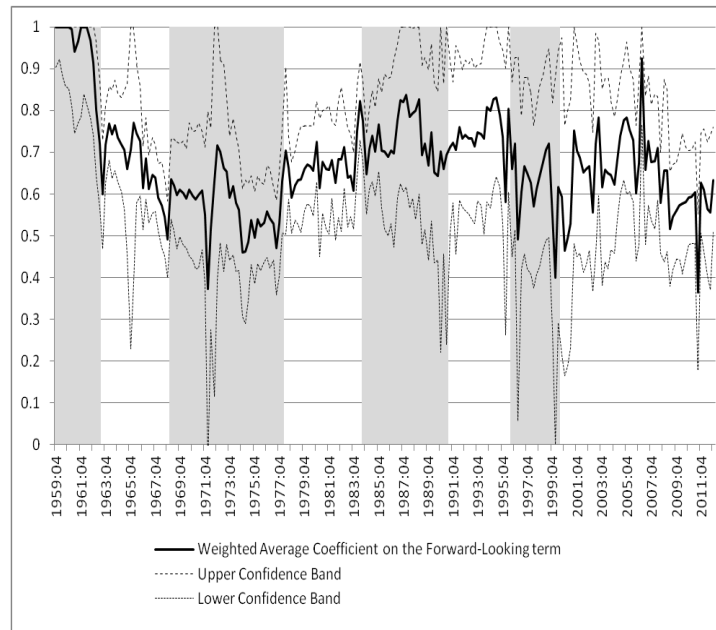


Figure D4.2: The evolution of the weighted average coefficient on the forward-looking term,  $\bar{\gamma}_{ft}$ , in the meta-Phillips Curve as generated by recursive GMM estimation, based on a higher probability of break.

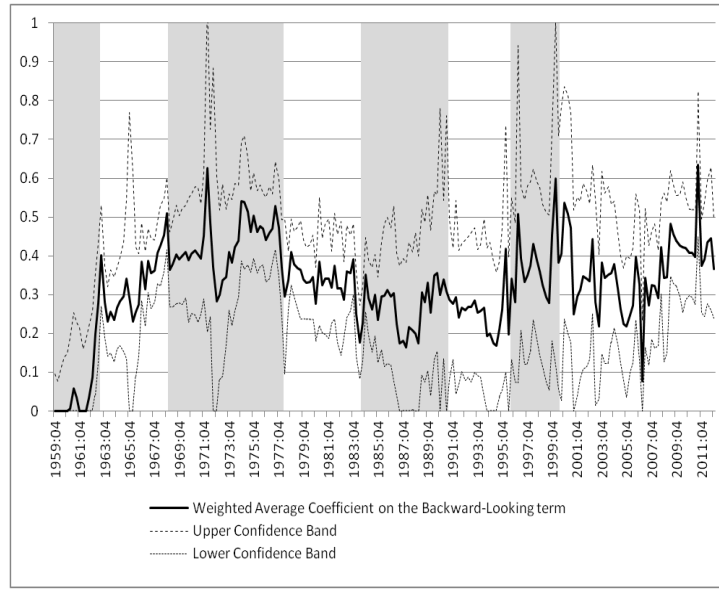


Figure D4.3: The evolution of the weighted average coefficient on the backward-looking term,  $\bar{\gamma}_{bt}$ , in the meta-Phillips Curve as generated by recursive GMM estimation, based on a higher probability of break.

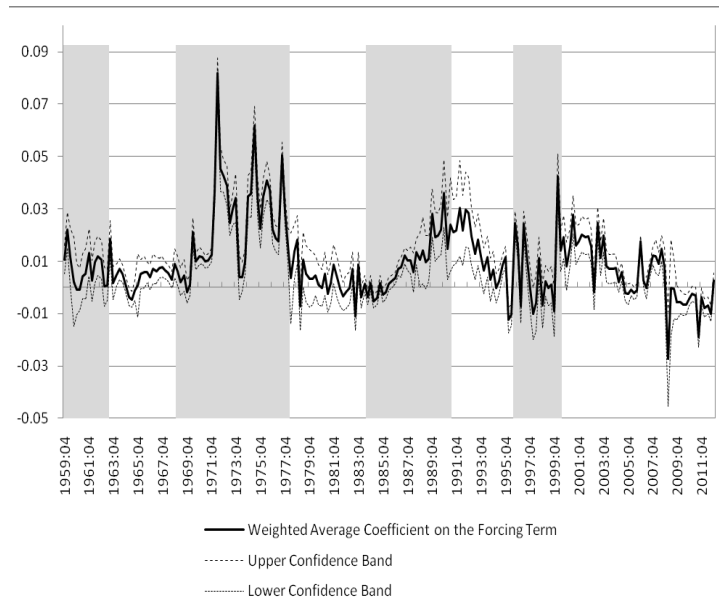


Figure D4.4: The evolution of the weighted average coefficient on the forcing variable,  $\bar{\lambda}_t$ , in the meta-Phillips Curve as generated by recursive GMM estimation, based on a higher probability of break.

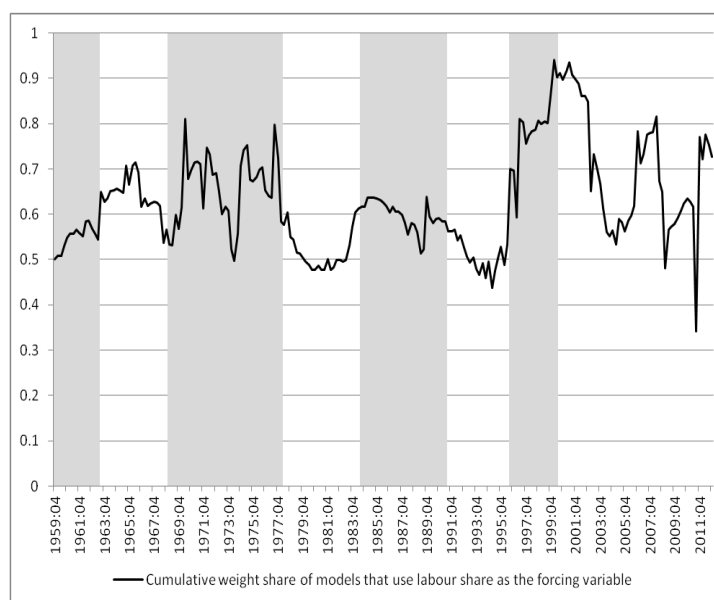


Figure D4.5: The evolution of weights allocated to models that use the labour share as the forcing variable, based on a higher probability of break.

## 4.10 References

1. Alogoskoufis, G.S. and R. Smith (1991), "The Phillips curve, the persistence of inflation, and the Lucas critique: evidence from exchange rate regimes," *American Economic Review*, 81, pp.1254-1275.
2. Atkeson, A. and L. E. Ohanian (2001), "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1), pp.2-11.
3. Ball, L., N. G. Mankiw and D. Romer (1988), "The New Keynesian Economics and the Output-Inflation Trade-off," *Brookings Papers on Economic Activity*, 19, pp.1-82.
4. Blanchard, O. and J. Galí. (2007), "The Macroeconomic Effects of Oil Price Shocks: Why Are the 2000s So Different from the 1970s?" *NBER Working Paper No. 13368*.
5. Bordo, M. D. and B. Eichengreen (2008), "Bretton Woods and the Great Inflation," *NBER-Working Paper No. w14532*. Available at SSRN: <http://ssrn.com/abstract=1312626>
6. Borio, C. E. V., and A. J. Filardo. (2007), "Globalisation and Inflation: New Cross-Country Evidence on the Global Determinants of Domestic Inflation," *BIS Working Paper No. 227*.

7. Calvo, G. A. (1983), "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, 12, pp.383-398.
8. Christiano, L J, M. Eichenbaum and C.L. Evans (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, pp.1-45.
9. Clarida, R., J. Galí and M. Gertler (1999), "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37, pp.1661-1707.
10. Clarida, R., J. Galí and M. Gertler (2000), "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115(1), pp.147-80.
11. Cogley, T. and T. J. Sargent (2001), "Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual*, pp.331-73.
12. Draper, D. (1995), "Assessment and Propagation of Model Uncertainty," *Journal of Royal Statistical Society Series B*, 58, pp.45-97.
13. Dufour, J.M. (1997), "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, pp.1365-89.
14. Erceg, C. J. and A.T. Levin (2003), "Imperfect credibility and inflation persistence," *Journal of Monetary Economics*, Elsevier, vol. 50(4), pp. 915-944.
15. Fuhrer, J. (1997), "The (Un)Importance of Forward-Looking Behavior in Price Specifications," *Journal of Money, Credit, and Banking*, 29, pp.338-350.
16. Fuhrer, J. and G. Moore (1995a), "Monetary policy trade-off and the correlation between nominal interest rates and real output," *American Economic Review*, 85(1), pp.219-239.
17. Fuhrer, J. and G. Moore (1995b), "Inflation Persistence," *Quarterly Journal of Economics*, 110, pp.127-159.
18. Fuhrer, J., G. Olivei, and G.M.B. Tootell (2009), "Empirical estimates of changing Inflation dynamics," *Federal Reserve Bank of Boston, Working paper*.

19. Gagnon, E. and H. Khan (2005), "New Phillips Curve under Alternative Production Technologies for Canada, the U.S., and the Euro Area," *European Economic Review* 49, pp.1571-1602.
20. Galí, J. and M. Gertler (1999), "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, 44, pp.195-222.
21. Galí, J., M. Gertler and D. Lopez-Salido (2001), "European inflation dynamics," *European Economic Review*, 45 (7), pp.1237-1270.
22. Galí, J., M. Gertler and D. Lopez-Salido (2005), "Robustness of Estimates of the Hybrid New Keynesian Phillips Curve," *Journal of Monetary Economics*, 52, pp.1107-1118.
23. Gaspar, V., F. Smets and D. Vestin (2006), "Adaptive learning, persistence and optimal monetary policy," *Journal of the European Economic Association*, Vol.4, Issue 2-3, pp.376-385.
24. Hoeting, J.A., D. Madigan, A.E. Raftery and C.T. Volinsky (1999), "Bayesian model averaging: A tutorial (with Discussion)," *Statistical Science*, 14, pp.382-401. Correction: vol. 15, pp.193-195.
25. Ihrig, J., S. B. Kamin, D. Lindner, and J. Marquez (2007), "Some Simple Tests of the Globalization and Inflation Hypothesis," *International Finance Discussion Paper No. 891*, Board of Governors of the Federal Reserve System.
26. Judd, J.P. and G. D. Rudebusch (1998), "Taylor's Rule and the Fed: 1970-1997," *Federal Reserve Bank of San Francisco Economic Review*, 3, pp.3-16.
27. Kim, C.J. and C.R. Nelson (1999), "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *Review of Economics and Statistics*, 81, pp.608-16.
28. Kim, C.J. and Y. Kim (2008), "Is the Backward-Looking Component Important in a New Keynesian Phillips Curve?" *Studies in Nonlinear Dynamics and Econometrics*, 12(3), pp.1-19.
29. Kleibergen, F. (2002), "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), pp.1781-1803.



30. Kleibergen, F. and S. Mavroeidis (2009), "Weak instrument robust tests in GMM and the new Keynesian Phillips curve," *Journal of Business and Economic Statistics*, 27, pp.293-311.
31. Lee, K., Morley J. and Shields K. (2015), "The meta Taylor Rule," *Journal of Money, Credit and Banking*, Vol.47, Issue 1, pp.73-98.
32. Levin, A. T. and J. M. Piger (2004), "Is Inflation Persistence Intrinsic in Industrial Economies?" European Central Bank Working Paper No. 334.
33. Ma, A. (2002), "GMM estimation of the new Phillips curve," *Economics Letters*, 76, pp.411-417.
34. Mavroeidis, S. (2005), "Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve," *Journal of Money, Credit and Banking*, 37, pp.421-48.
35. Melick, W. and G. Galati (2006), "The Evolving Inflation Process: An Overview." BIS Working Paper No. 196.
36. Mishkin, F.S. (2007), "Inflation Dynamics", *International Finance*, Vol.10, Issue 3, pp.317-334.
37. Musso, A., L. Stracca and D. van Dick (2009), "Instability and Nonlinearity in the Euro-Area Phillips Curve," *International Journal of Central Banking*, Vol. 5 No. 2, pp.181-212.
38. Neiss, K.S. and E. Nelson (2002), "Inflation Dynamics, Marginal Cost, and the Output Gap: Evidence from Three Countries," Working Paper, Bank of England.
39. Orphanides, A. and J.C. Williams (2006), "Monetary Policy with Imperfect Knowledge," *Journal of the European Economic Association*, MIT Press, vol. 4(2-3), pp.366-375.
40. Orphanides, A. and J.C. Williams (2008), "Learning, Expectations Formation, and the Pitfalls of Optimal Control Monetary Policy," *Journal of Monetary Economics*, 55, pp.80–96.
41. Orphanides, A. and J.C. Williams (2012), "Monetary Policy mistakes and the Evolution of Inflation Expectations", a chapter in *The Great Inflation*, from National Bureau of Economic

Research, Inc.

42. Pesaran, M.H. (1987), "The Limits to Rational Expectations," Oxford: Blackwell Publishers.
43. Raftery, A.E., D. Madigan and A. Hoeting (1997), "Bayesian Model Averaging for Regression Models," *Journal of the American Statistical Association*, 92, pp.179-191.
44. Roberts, J. M. (1995), "New Keynesian Economics and the Phillips Curve," *Journal of Money, Credit, and Banking*, 27, 4, part 1, November, pp.975-984.
45. Roberts, J. M. (1998), "Inflation Expectations and the Transmission of Monetary Policy," Finance and Economics Discussion Series Paper No. 43, Board of Governors of the Federal Reserve System.
46. Roberts, J. M. (2001), "How well does the New Keynesian Sticky-price model fit the data?" Board of Governors of the Federal Reserve System.
47. Roberts, J. M. (2006), "Monetary Policy and Inflation Dynamics." *International Journal of Central Banking*, September, pp. 193-230.
48. Rotemberg, J. (1982), "Monopolistic Price Adjustment and Aggregate Output," *The Review of Economic Studies*, 49, pp.517-531.
49. Rotemberg, J. and M. Woodford (1997), "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Annual 1997*, MIT Press.
50. Rudd, J. and K. Whelan (2005a), "Does Labor's Share Drive Inflation?" *Journal of Money, Credit, and Banking*, 37, pp.297-312.
51. Rudd, J. and K. Whelan (2005b), "New Tests of the New-Keynesian Phillips Curve," *Journal of Monetary Economics*, 52, pp.1167-1181.
52. Rudd, J. and K. Whelan (2006), "Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?" *American Economic Review*, 96, 1, pp.303-320.

53. Rudd, J. and K. Whelan (2007), "Modelling Inflation Dynamics: A critical Review of Recent Research," *Journal of Money, Credit and Banking*, 39, s1, pp.155-170.
54. Russell, B, A. Banerjee, I. Malki and N. Ponomareva, (2010), "A Multiple Break Panel Approach To Estimating United States Phillips Curves," *Dundee Discussion Papers in Economics* 232, *Economic Studies*, University of Dundee.
55. Sbordone, A. (2002), "Prices and Unit Labor Costs: A New Test of Price Stickiness," *Journal of Monetary Economics*, 49(2), pp.265-292.
56. Sbordone, A. (2005), "Do Expected Future Marginal Costs Drive Inflation Dynamics?" *Journal of Monetary Economics*, 52(6), pp.1183-1197.
57. Staiger, D. and J.H. Stock (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica* 65(3), pp.557-86.
58. Stock, J.H., J.H. Wright and M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics* 20(4), pp.518-29.
59. Taylor, J. B. (2000), "Low Inflation, Pass-through, and the Pricing Power of Firms," *European Economic Review*, 44, pp.1389-408.
60. Taylor, J.B. (1980), "Aggregate dynamics and staggered contracts," *Journal of Political Economy* 88, pp.1-23.
61. Taylor, J.B. (1999), "A Historical Analysis of Monetary Policy Rules," in John B. Taylor (ed.) *Monetary Policy Rules*, University of Chicago Press.
62. Willis, J. L. (2003), "Implications of Structural Changes in the U.S. Economy for Pricing Behavior and Inflation Dynamics," *Federal Reserve Bank of Kansas City Economic Review*, Q1, pp.5-27.
63. Zhang, C., D.R. Osborn and D.H. Kim (2008), "The New Keynesian Phillips Curve: From Sticky Inflation to Sticky Prices," *Journal of Money, Credit and Banking*, Vol. 40, No. 4, pp.667-699.

## Chapter 5

# Nowcast and Forecast Evaluation when Real-Time Data are Available

### 5.1 Introduction

The use of real-time datasets has become increasingly popular since they can shed light on how policymakers design policies based on information that is available at the time decisions are taken. The datasets contain the available history of data vintages, showing the preliminary estimates of variables at the earliest opportunity alongside their subsequent revision as more complete information becomes available. The datasets also often include direct measures of expectations as expressed in surveys published at the time, describing agents' beliefs on expected future values of the macroeconomic variables and the expected contemporaneous value when the first-release data are published with a delay. In recent years, real-time datasets have become the forefront of macroeconomic analysis as researchers extend the methods of using such datasets in prescribing and evaluating policy (see, for example, Croushore (2011) for recent developments).

One area in which real-time data are potentially important is in forecasting since the data provide a comprehensive description of the context in which forecasts and subsequent decisions are made. However, while real-time data are employed in forecasting, there remains some scepticism about their usefulness and they do not figure in forecasting exercises as often as might be expected. The purpose of this chapter then is to judge the usefulness of real-time datasets in decision-making, focusing on the importance of using revisions data and survey data in nowcasting and forecasting quarterly output growth and price inflation in the US.

The contribution to literature is threefold. *First*, we propose a simple canonical modelling

framework that can readily accommodate revisions and survey data alongside the most recent data measures to characterise the underlying data generating process of the variables of interest as well as the expectation formation and measurement processes. *Secondly*, we investigate the importance of revisions and survey data by comparing the nowcasting and forecasting performance of models that fully make use of the data compared to the performance of models that are based only on the most recent vintage of data or which make only partial use of the information contained in real-time datasets. To enable this, we adopt model averaging methods in a recursive set-up, accommodating the possibility that the usefulness of real-time data changes over time (diversifying against possible structural breaks). A robustness exercise explicitly allows for structural breaks by averaging across different estimation windows, gaining important insights about the trade-off between achieving more accurate parameter estimates and delivering forecasts that reflect the existence of breaks. *Thirdly*, the chapter considers a range of evaluation criteria to judge the usefulness of real-time data in forecasting, drawing an important distinction between ‘real-time forecast evaluation’ and a ‘final assessment’ of forecast performance. We judge the usefulness of real time data according to statistical criteria that distinguish between the performance of point forecasts and density forecasts and we also consider the forecast performance according to economic criteria. In the latter case, we focus on a set of events involving output and inflation that are frequently discussed in business cycle analysis and introduce a novel means of evaluating these probability forecasts, based on fair-bet, to investigate the role of real-time data in forecast-based decisions involving relatively rare events.

The three aspects of our modelling exercise are motivated by different strands of the literature. The incorporation of both survey information and data revisions in real-time data analysis has been widely examined by a number of researchers who often rely on mixed-frequency data, dealing with unsynchronised inflow of data as the nowcast of a series of interest, such as quarterly output, is almost continuously updated in the light of new information (Giannone, Reichlin and Small (2008), Arouba et al. (2009), Ghysels and Wright (2009), and Faust and Wright (2009)). However, whether survey and revisions data can become useful to the researchers has been viewed with scepticism. As far as revisions data are concerned, Mankiw and Shapiro (1986) were among the first to conclude that revisions contain unforecastable new information and are certainly not noise, indicating that there is little scope for using other observed data to improve the estimate of the underlying series.

Croushore and Stark (2003) and Croushore and Evans (2006) demonstrate the problems associated with the characterisation of the revision process while Koenig, Dolmas and Piger (2003) argue that because researchers tend to mix heavily revised data with first-release and lightly revised data, the conventional approach of using all types of data is unlikely to yield good forecasts. As they demonstrate, under certain circumstances first-release data should be preferred to real-time data since the linkages between observed data and forecasts are likely to be different at the start of the sample period, where data are heavily revised. The authors emphasise the biases arising in standard analysis of data subject to revisions, where the most recent measures of the variables are used in forecasting models. Patterson (2002), Kishor and Koenig (2005), Garratt et al. (2008), Clements and Galvão (2013) and Jacobs and van Norden (2011) all argue that revisions contain useful information.

There has not been a consensus on the usefulness of survey data either. Some authors support the view that surveys can contain important information held at the time of implementation of decisions which is not captured by the available measures of actual variables. In other words, surveys reflect the real time use of information and can therefore improve models' nowcast and forecast performance. Croushore (2010), Ang et al. (2007) and Aretz and Peel (2010) show that survey expectations are often hard to beat in real time forecasting exercises. Bańbura and Rünstler (2011) demonstrate the need to use survey data in nowcasts based on mixed-frequency modes while Matheson et al. (2010) argue that survey data are particularly helpful when predicting actual series and their revisions.

The second aspect of the chapter focuses on the usefulness of revisions and survey data in forecasting and relates to the use of information when there are many potential predictor variables, as discussed in Clements and Hendry (2005) and Stock and Watson (2006) for example. This literature recognises that, with the samples of data typically available, parameter estimation error can dominate model's forecast performance. This means, for example, that adding a variable to a forecasting model can undermine its forecasting performance even if the variable is part of the true data generating process. One way to mitigate against this problem is to average across forecasts from different models (see, for example, Harvey and Newbold (2005) and Timmermann (2006) for discussion). The idea is that if two models are misspecified and provide incomplete information

about the true data generating process, then some combination of the two could potentially reduce the bias generated by the individual models themselves. This is the approach taken here, producing forecasts using various ‘average’ models each constructed using model averaging techniques. The average models are distinguished according to their use of the real time data (making use of vintage data only, survey data only or both). The averaging allows for time-varying weights and ensures that each average model makes best use of the information available to it in forecasting. Comparison of the forecasts across the average models then provides an assessment of the contributions of the different types of real-time data.

The third aspect of the chapter relates to the ambiguity on the criteria to be used in forecast evaluation. This partly arises from an increasing awareness of the importance of properly characterising forecast uncertainties which has shifted attention from point forecasts to density forecasts, and evaluation criteria from models’ root mean squared errors (RMSEs) to their probability integral transforms (PITs) and logarithmic scores; see, for example, the June 2010 Special Issue of *Journal of Applied Econometrics* for an overview. But there is also increasing interest in judging the economic value of a model’s forecast, concentrating on the usefulness of the models in a specific decision-making context rather than on its statistical performance, as discussed in Granger and Machina (2006) for example. Certainly economic and statistical evaluation criteria highlight different features of the models and their forecast performance and so, in this chapter, we judge the usefulness of real time data not just in terms of their use in generating point and density forecasts of output growth but also their role in forecasting the probability of relatively rare/extreme recessionary events.

The layout of the remainder of the chapter is as follows<sup>1</sup>. Section 5.2 carefully describes our modelling framework, distinguishing between the real-time assessment in which model weights are determined, and the final assessment in which the out-of-sample forecast performance of the various average models is compared. Moreover, we explain how the use of density and probability forecasts can be used in model evaluation based on both statistical and economic criteria. The evaluation is performed by considering a range of events around a fair bet. Section 5.3 applies the methods

---

<sup>1</sup>The work in sections 5.1-5.3 of this chapter is based on Aristidou, Lee and Shields (2015).

to US data over 1968q4 – 2014q3, including all the data vintages available for actual and expected output and price data from surveys over the period. We conclude that judged by statistical criteria, the performance of models that nowcast and forecast output growth and inflation is considerably enhanced by taking into account information available in survey data and revisions data. We also show that relating to economic significance, both revisions data and survey expectations are important in calculating density forecasts in forecasting the occurrence of a number of events related to the real business cycle. Moreover, the use of revisions and survey data can substantially boost investor's profits in decision-making based on forecasts, as shown by a "fair bet" exercise. In addition, the analysis demonstrates that evaluating forecasts based on future growth and inflation dynamics that are relevant to decision-makers is a much more reliable practice than relying solely on point forecasts. In section 5.4 we consider an extension of our methodological framework to explicitly allow for the possibility of structural breaks in the relationships of interest. The real and final assessments are repeated but we also average across different sample sizes over which individual models hold. Results on this approach are presented in section 5.4.2. As it turns out, we show that averaging across different sample sizes to address structural breaks does not change results significantly with the weight structure assigning more weights to models estimated over longer sample periods. Section 5.5 concludes with a brief summary of the findings.

## 5.2 Assessing the Usefulness of Real-Time Information

### 5.2.1 A Modelling Framework to Accommodate Real-Time Information

In what follows, we write the (logarithm of the) variable  $x$  at time  $t - s$  as  $x_{t-s}$ , and denote the measure of variable  $x$  at time  $t - s$  that is released at time  $t$  by  ${}_tx_{t-s}$ ,  $s = 0, 1, 2, \dots$ , while  ${}_tx_{t+s}^e$  is a direct measure of the expected value of the variable at  $t + s$ , with the expectation formed on the basis of information available at the time the measure is released,  $t$ . Here we consider more than one variable (output,  ${}_ty_{t-s}$ , and price inflation,  ${}_tp_{t-s}$ ) so we write  ${}_t\mathbf{x}_{t-s} = ({}_ty_{t-s}, {}_tp_{t-s})'$ , a  $2 \times 1$  vector of variables. Throughout, we shall assume that data is published with a one period delay, and the time- $t$  vintage of data is denoted  $\mathbf{X}_t = \{{}_t\mathbf{x}_1, {}_t\mathbf{x}_2, \dots, {}_t\mathbf{x}_{t-2}, {}_t\mathbf{x}_{t-1}, {}_t\mathbf{x}_t^e, {}_t\mathbf{x}_{t+1}^e, \dots, {}_t\mathbf{x}_{t+F}^e\}$  which includes the time- $t$  measures of the actual variables at  $t = 1, \dots, t - 1$  and the time- $t$  measures of



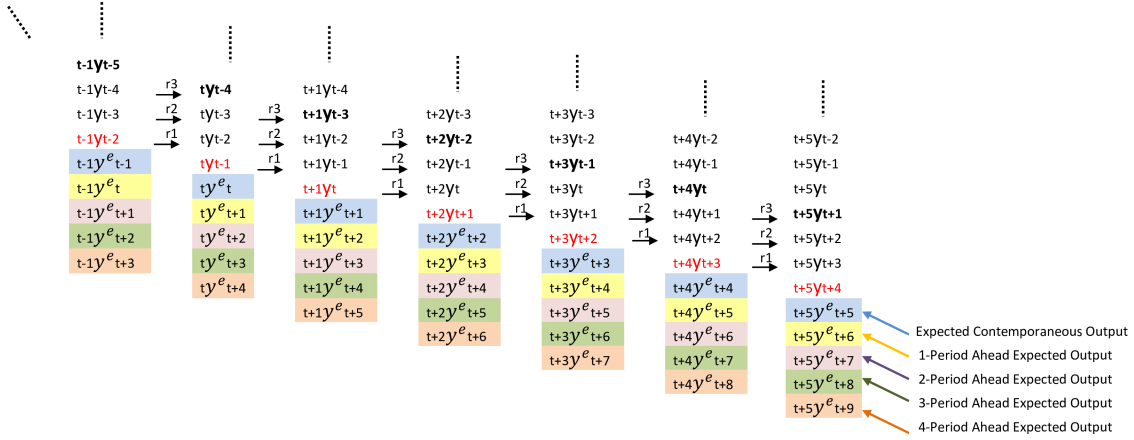


Figure 5-1: The form of the real time data set employed. Figures in bold denote the actual post-revision series, assuming revisions continue no longer than three periods, the red figures denote the first release output series. The shaded figures denote information from surveys.

expected contemporaneous and future values of the variables published for up to  $F$  periods ahead. In our real-time forecast evaluation exercise, we denote the period in which decisions are made by  $\tau$  for  $\tau \leq T$  and  $\mathbf{X}_\tau$  is termed ‘the most recent data vintage’ while  $\mathbf{X}_T$  is the ‘final data vintage’. The information set grows with the addition of successive vintages of datasets by including the:

- news on the variables in the previous period (the "first release" of information on the output and price level in that period);
- news on the revisions of the output series in the previous periods, i.e.  $\{(t+1y_{t-1}-ty_{t-1}), (t+1y_{t-2}-ty_{t-2}), \dots\}$  and
- news on the expectations of the contemporaneous and future values of the variables.

Figure 5.1 illustrates the form of the real time data set employed. Here, we assume that revisions continue no longer than three periods. Accordingly, the first release information on output at  $t$  i.e.  $t+1y_t$ , is revised three times so that the final post-revision output series is given by  $t+4y_t$ . Figure 5.1 also demonstrates the use of survey information. Since data is published with one period delay, agents use information from surveys for the contemporaneous realisation of the output series, as

well as for future values of the output series. Here we assume that the survey horizon extends no longer than four periods ahead.

For the modelling exercise, we assume that the variables  $x_{1t}, \dots, x_{mt}$  are all difference stationary, that revisions are stationary and that expectational errors are stationary. If revisions continue for no longer than  $R$  periods after the first-release and if the surveys provide measures of the expected values of  $\mathbf{x}$  for up to  $F$  periods ahead, then in making a decision at time  $\tau$  we will require a model that explains the following  $n = 1 + R + F + 2$  series over the period  $t = 1, \dots, \tau$ :

$$\begin{aligned} \mathbf{D}\mathbf{x}_t &= {}_{\tau}\mathbf{x}_{t-1} - {}_{\tau}\mathbf{x}_{t-2} : \text{growth in the most-recent data vintage for } \mathbf{x}; \\ \mathbf{D}\mathbf{x}_{rt} &= {}_t\mathbf{x}_{t-1-r} - {}_{t-1}\mathbf{x}_{t-1-r} : r^{th} \text{ revision of } \mathbf{x}, r = 0, \dots, R; \\ \mathbf{D}\mathbf{x}_{ft} &= {}_t\mathbf{x}_{t+f}^e - {}_t\mathbf{x}_{t+f-1}^e : \text{expected growth of } \mathbf{x} \text{ from surveys, } f = -1, 0, \dots, F. \end{aligned} \quad (5.1)$$

A simple  $p$ -order vector autoregressive model that explains these series for  $t = 1, \dots, \tau$  can be written:

$$\mathbf{D}\mathbf{x}_t = \mathbf{A}_{10} + \sum_{i=1}^p \left[ \mathbf{A}_{11i} \mathbf{D}\mathbf{x}_{t-i} + \sum_{r=1}^R \mathbf{A}_{12ri} \mathbf{D}\mathbf{x}_{r,t-i} + \sum_{f=1}^F \mathbf{A}_{13fi} \mathbf{D}\mathbf{x}_{f,t-i} \right] + \boldsymbol{\varepsilon}_{1t}, \quad (5.2)$$

$$\begin{aligned} \mathbf{D}\mathbf{x}_{r,t} &= \mathbf{A}_{20} + \sum_{i=1}^p \left[ \mathbf{A}_{21i} \mathbf{D}\mathbf{x}_{t-i} + \sum_{r=1}^R \mathbf{A}_{22ri} \mathbf{D}\mathbf{x}_{r,t-i} + \sum_{f=1}^F \mathbf{A}_{23fi} \mathbf{D}\mathbf{x}_{f,t-i} \right] + \boldsymbol{\varepsilon}_{2rt} \\ r &= 0, \dots, R \end{aligned} \quad (5.3)$$

$$\begin{aligned} \mathbf{D}\mathbf{x}_{f,t} &= \mathbf{A}_{30} + \sum_{i=1}^p \left[ \mathbf{A}_{31i} \mathbf{D}\mathbf{x}_{t-i} + \sum_{r=1}^R \mathbf{A}_{32ri} \mathbf{D}\mathbf{x}_{r,t-i} + \sum_{f=1}^F \mathbf{A}_{33fi} \mathbf{D}\mathbf{x}_{f,t-i} \right] + \boldsymbol{\varepsilon}_{3ft} \\ f &= -1, 0, \dots, F \end{aligned} \quad (5.4)$$

where the  $\mathbf{A}$ 's are matrices of coefficients and  $\boldsymbol{\varepsilon}$ 's are vectors of mean-zero shocks. We denote this model by  $M_{R,F,\tau}$  in what follows, with the subscript ' $R, F, \tau$ ' highlighting that the estimated model will differ depending on the underlying assumptions on the maximum number of revisions, the forecast horizon in the survey and on the estimation period.

Noting that the variables in  $\mathbf{D}\mathbf{x}_t$ ,  $\mathbf{D}\mathbf{x}_{r,t}$ , and  $\mathbf{D}\mathbf{x}_{f,t}$  involve the  $\mathbf{x}$  variables measured at  $t$  and earlier, the equations in (5.2)-(5.4) can be stacked and transformed to obtain the  $(p+1)^{th}$ -order

autoregressive model

$$\mathbf{z}_t = \mathbf{B}_0 + \sum_{i=1}^{p+1} \mathbf{B}_{1i} \mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, \tau \quad (5.5)$$

where  $\mathbf{z}_t = ({}_t\mathbf{x}_{t-1}, {}_t\mathbf{x}_{t-2}, \dots, {}_t\mathbf{x}_{t-R+1}, {}_t\mathbf{x}_t^e, \dots, {}_t\mathbf{x}_{t+F}^e)'$ , and  $\mathbf{B}_0$  is a  $mn \times 1$  vector of parameters derived from the  $\mathbf{A}$ 's and the  $\mathbf{B}_{1i}$  are  $mn \times mn$  matrices similarly derived. The structure of the model in (5.2)-(5.4) means that it can be rewritten in the form of a cointegrating VAR in which the parameters are restricted to reflect the assumptions that revisions and expectational errors are stationary.<sup>2</sup> The cointegrating VAR can then be rewritten in the form at (5.5) which is more convenient for describing simulation and forecasting exercises but which will retain the property that the various measures of the variables in  $\mathbf{x}$  converge to the same values in the long run.

It is worth noting that the above transformation incorporates the assumption that the data is revised no more than  $R$  times. Taken literally, this means that the most recent vintage of data  $\mathbf{x}_\tau$  describes the post-revision series for observations dated at  $\tau - R$  and earlier (i.e.  ${}_t\mathbf{x}_{\tau-s} = {}_\tau\mathbf{x}_{\tau-s}$ ,  $t = 1, \dots, \tau$ ,  $s = R + 1, \dots$ ). If model (5.5) is estimated in real-time based only on the most recent vintage of data, so that it does not account for either past vintages of data or the expectations of the contemporaneous or future values of  $\mathbf{x}$  as provided in surveys, then is called the ‘*conventional model*’ or ‘*quasi model*,’ denoted by  $M_{0,-1,\tau}$ .

### 5.2.2 Assessing Nowcasting/Forecasting Performance

In judging the usefulness of real-time data in decision-making, we shall assume that decisions in time  $\tau$  require a nowcast to be made of the *true* current state of the economy which we take to be measured by the post-revision measure  ${}_{\tau+R+1}\mathbf{x}_\tau$ . Our assessment of the usefulness of real-time data in decision-making is based on the nowcasting performance of decision-making strategies that make full use of the revisions and survey data compared to simpler strategies that make only partial use of the data (including the conventional real-time modelling approach, for example). It is worth providing a brief overview of the approach to assessment before formalising it in detail below. Our approach includes two complementary elements: a real-time assessment and a final assessment.

---

<sup>2</sup>This means that there are  $R$  cointegrating relations between  ${}_ty_{t-1}$  and each of the  ${}_t\mathbf{x}_{t-1-R}$ , and  $F+1$  cointegrating relations between  ${}_t\mathbf{x}_{t-1}$  and each of the  ${}_t\mathbf{x}_{t+F}^e$ , all of the form  $(1, -1)$ . See Garratt et al. (2008) for details.

The *real-time assessment* is based on the estimation of an ‘average model’ obtained using model averaging methods applied to models that are all estimated on data available at the time but are distinguished by their different use of the survey and revisions data. The use of model averaging methods means the advantages of forecast combinations are exploited and the value of the real-time data is captured properly. Further, the weights given to the different models are chosen to reflect their relative nowcasting performance so that they provide a straightforward summary of the usefulness of the revision and survey data in nowcasting as it would be judged in real time.

The real-time assessment of models could change over time since survey and revisions data might be more or less useful in different circumstances. It is therefore useful to consider the extent to which the real-time data is helpful in strategic decision-making judged over a longer evaluation period therefore. The average model obtained following the steps described above would form the basis of decisions made in any given period (each time making use of all the revisions and survey data available). The nowcasting performance of the average models obtained in this way for each period in our sample of data, up to and including the final vintage, can be used to judge the usefulness of the real-time data over the whole sample. The *final assessment* obtained in this way can be compared to equivalent assessments based on an alternative set of average models obtained from analyses in which survey and/or revisions data are not used. This comparison judges the overall usefulness of real-time data, asking whether it is a good idea to include real-time data in the forecasting exercise generally even though, at particular times, it might not appear useful.

In terms of the criteria to be used, it is worth noting that recent years have seen a growing interest in a decision-based approach to the evaluation and comparison of forecasts. Here, forecast accuracy is judged according to its economic value given an explicitly defined decision-making context. This reflects the recognition that the statistical criteria used to evaluate models, are typically based solely around point forecasts and measured using mean squared forecasting error (MSE), provide information on the economic value of their forecasts only under certain conditions. The preponderance of studies employing the decision-based approach to forecast evaluation are in the area of applied finance where investment strategies and their outcomes are relatively

straightforward to describe and measure.<sup>3</sup> The decision-making context in macroeconomics is not so straightforward and there is no generally accepted decision-based criterion with which to judge models' forecasts of output and inflation fluctuations. However, a comprehensive statement on the usefulness of real time data in forecasts of these variables should include an evaluation element that reflects the variety of ways in which recession and business cycle fluctuations are experienced by different individuals so that the economic worth of the forecast can be judged in addition to using statistical criterion.

### **Real-time assessment**

The VAR model of (5.2)-(5.5) provides a simple framework within which all the real-time data available can be accommodated in a coherent way. But such a model could be very highly parameterised, depending on the number of revisions available, the length of the survey horizon and the chosen order of the VAR, and this could undermine its value as a forecasting tool. Similarly, it is possible that different parts of the real-time data become more or less useful for forecasting at different times. For example, statistical agencies' procedures could mean that measurement errors contained in the first-release of data are more pronounced in times of very high or very low growth, making revisions data more useful. Or forecasters may watch incoming news more carefully at times of crisis so that survey data becomes more informative at these times.<sup>4</sup>

To mitigate against these problems, we estimate a set of models of the form in (5.2)-(5.5) (distinguished by their use of real-time information) and combine these using model averaging techniques into an average model. The weights used to combine the models, and their associated forecasts, can be chosen so that forecast performance is maximised (avoiding the problems of over-parameterisation) and can change over time so that different parts of the real-time data can be used when they are helpful. The approach exploits the fact that forecast performance is improved through a linear combination, as established by Harvey and Newbold (2005).

---

<sup>3</sup>See, for example, Leitch and Tanner (1991), Abhyankar et al. (2005) and Garratt and Lee (2010).

<sup>4</sup>See, Loungani et al. (2013), for example, for discussion on the changing impact of information rigidities on survey data at different points of the business cycle.

A strategy that makes full use of the data could base its decisions at  $\tau$  on forecasts from model  $M_{R,F,\tau}$  or indeed any model  $M_{r,f,\tau}$  for  $r = 0, \dots, R$  and  $f = -1, \dots, F$ , or a weighted average of these models in what we term an ‘average model’. The weights can be chosen to reflect the relative forecasting performance over the recent past, where performance is judged according to the question of interest. So here, the weights are based on the models’  $(R + 1)$ -period-ahead forecasts of the (post-revision) measure of output,  $\tau_{+R+1}y_\tau$ , if the ‘true’ variable level is of interest. Then the average model that makes full use of the real-time data is obtained as follows:

- Split the currently available sample into two sub-samples: an estimation period  $t = 1, \dots, \tau - \Lambda$ ; and a ‘training period’  $t = \tau - \Lambda + 1, \dots, \tau$ .
- Estimate model defined in (5.2)-(5.4) over  $t = 1, \dots, \tau - \Lambda$  and for  $r = 0, \dots, R$  and  $f = -1, \dots, F$ , providing  $(R + 1) \times (F + 2)$  alternative models.
- Evaluate the forecast performance of these individual models by comparing the nowcasts based on the models,  $E[\tau_{-\Lambda+1+R}y_{\tau-\Lambda} \mid M_{r,f,\tau-\Lambda}]$ , with the observed post-revision value  $\tau_{-\Lambda+1+R}y_{\tau-\Lambda}$  through the calculation of a chosen statistical criterion. If the interest is primarily on the performance of models in terms of point forecasts, then the squared forecast error  $s_{r,f,\tau-\Lambda}^p = (\tau_{-\Lambda+1+R}y_{\tau-\Lambda} - E[\tau_{-\Lambda+1+R}y_{\tau-\Lambda} \mid M_{r,f,\tau-\Lambda}])^2$  is used. If the whole density forecast is of interest, we use the logarithmic score  $s_{r,f,\tau-\Lambda}^d = \ln(g(\tau_{-\Lambda+1+R}y_{\tau-\Lambda} \mid M_{r,f,\tau-\Lambda}))$  where  $g(\tau_{-\Lambda+1+R}y_{\tau-\Lambda} \mid M_{r,f,\tau-\Lambda})$  is the nowcast density for model  $M_{r,f,\tau-\Lambda}$ .
- Repeat exercise for samples over the whole of the training period. So, for instance, the next  $(R + 1) \times (F + 2)$  versions of models are estimated over  $t = 1, \dots, \tau - \Lambda + 1$ , and the nowcast performance of model  $M_{r,f,\tau-\Lambda+1}$  is based on  $E[\tau_{-\Lambda+2+R}y_{\tau-\Lambda+1} \mid M_{r,f,\tau-\Lambda+1}]$ , producing the squared forecast error,  $s_{r,f,\tau-\Lambda+1}^p$ , and logarithmic score,  $s_{r,f,\tau-\Lambda+1}^d$ . The process continues until the final versions of model  $M_{r,f,\tau}$  are estimated producing the corresponding  $s_{r,f,\tau}^p$  and  $s_{r,f,\tau}^d$ .
- Calculate weights for the models,  $w_{r,f,\tau}$ , on the basis of the relative forecast performance of the individual models over the training period, using the average mean squared forecast error

for each model over the training period (i.e.  $MSE_{r,f,\tau} = \frac{1}{\Lambda} \sum_{\lambda=0}^{\Lambda} s_{r,f,\tau-\Lambda+\lambda}^p$ ), as follows:

$$w_{r,f,\tau} = \frac{(\sqrt{MSE_{r,f,\tau}})^{-1}}{\sum_r \sum_f (\sqrt{MSE_{r,f,\tau}})^{-1}} \quad (5.6)$$

or the average of the logarithmic score over the training period, (i.e.  $MLS_{r,f,\tau} = \frac{1}{\Lambda} \sum_{\lambda=0}^{\Lambda} s_{r,f,\tau-\Lambda+\lambda}^d$ ) and weights are given by:

$$w_{r,f,\tau} = \exp(MLS_{r,f,\tau}) / \sum_r \sum_f \exp(MLS_{r,f,\tau}) \quad (5.7)$$

The ‘average model’ that makes full use of the real-time data over the period  $t = 1, \dots, \tau$  consists of the the individual estimated models and their weights and it is denoted by:

$$\overline{M}_{R,F,\tau} = \{M_{r,f,\tau}, w_{r,f,\tau} \text{ for } r = 0, \dots, R \text{ and } f = -1, \dots, F\}. \quad (5.8)$$

The average model can be used to obtain point forecasts and density forecasts using the weighted average of the models’ individual point forecasts and aggregating over the models’ individual densities.

The nature of the average model obtained at time  $\tau$  can be summarised by the statistics

$$\mu_{\tau}^r = \sum_{r=0}^R r \times w_{r,f,\tau} \quad \text{and} \quad \mu_{\tau}^f = \sum_{f=-1}^F f \times w_{r,f,\tau}. \quad (5.9)$$

The  $\mu_{\tau}^r$  and  $\mu_{\tau}^f$  statistics capture the relative importance of the revision data and the survey data in defining the average model at time  $\tau$ . If they deviate from 0 and  $-1$  respectively, they show that the revisions data and the survey data would have made a contribution to an out-of-sample forecasting exercise if it had been conducted in real time. This provides a *real-time assessment* of the usefulness of the revisions and survey data in  $\overline{M}_{R,F,\tau}$ . Different values for the weights could be obtained for different  $\tau$ , allowing the possibility that the usefulness of the revision and survey data changes over time.

### Final-time assessment

The average model  $\overline{M}_{R,F,\tau}$  can be used to provide the nowcast of  $\mathbf{x}_{\tau}$  for any decision-date  $\tau$ . If this exercise is repeated over an evaluation period ( $\tau = \underline{\tau}, \dots, T$ , say), then the performance criterion

can be calculated in each exercise judged from the perspective of the whole sample of data up to and including the final vintage. The overall performance over the evaluation period provides a *final assessment* of the usefulness of the decision-making strategy which makes full use of the available revisions and survey data and fully exploits the potential contribution the real time data can make to forecasting. Corresponding exercises could also be undertaken to obtain a final assessment of strategies where

- no use is made of the survey data throughout (i.e. based on the average model  $\overline{M}_{R,-1,\tau}$  for  $\tau = \underline{\tau}, \dots, T$ );
- no use is made of the revisions data throughout (i.e. based on the average model  $\overline{M}_{0,F,\tau}$  for  $\tau = \underline{\tau}, \dots, T$ ); and
- no use is made of either the revisions or survey data throughout (i.e. based on the conventional real time model  $\overline{M}_{0,-1,\tau}$  for  $\tau = \underline{\tau}, \dots, T$ ).

These three models are nested within  $\overline{M}_{R,F,\tau}$ , and, in principle, could be chosen if zero weights are placed on the models involving revisions or surveys at all times when deriving  $\overline{M}_{R,F,\tau}$ . In practice, zero weights might be unlikely and so comparison of the forecast criteria obtained from the four models provides an overall assessment of the usefulness of the revisions and survey data taking into account that they might be more or less useful at different times.

**Statistical forecast evaluation criteria** The evaluation of nowcasts/forecasts obtained from average models can be based on statistical criteria such as MSE or logarithmic score. The MSE concentrates on the point forecast of the post-revision series and, in the final assessment, it is measured by  $\sum_{\tau=\underline{\tau}}^T (\tau_{+1+R}y_{\tau} - E[\tau_{+1+R}y_{\tau} | \overline{M}_{R,F,\tau}])^2$  in the case of the average model in which all the real time data is used.<sup>5</sup> Of course, at the point at which the final assessment of the average model is made, the weights have already been calculated. If the final assessment is based on the MSE, then the weights underlying the average model should also be based on the forecast

---

<sup>5</sup>  $E[\tau_{+1+R}\mathbf{x}_{\tau} | \overline{M}_{R,F,\tau}]$  is replaced by  $E[\tau_{+1+R}\mathbf{x}_{\tau} | \overline{M}_{R,-1,\tau}]$  in the case of the model where no survey data is used; by  $E[\tau\mathbf{x}_{\tau} | \overline{M}_{0,F,\tau}]$  in the model where no survey data is used; and by  $E[\tau\mathbf{x}_{\tau} | \overline{M}_{0,-1,\tau}]$  in the conventional model.



performance of the individual models judged by the MSE over the training period; e.g. which gives a high weight to models with relatively low MSE's.

The logarithmic score is based on the combined forecast density defined, in the case of the average model  $\bar{M}_{R,F,\tau}$  for example, by  $p(y_\tau | \bar{M}_{R,F,\tau}) = \sum_r \sum_f w_{r,f,\tau} g(y_\tau | M_{r,f,\tau})$  where  $g(y_\tau | M_{r,f,\tau})$  is the nowcast density for model  $M_{r,f,\tau}$ . The logarithmic score at  $\tau$  evaluates the nowcast density at the post-revision value, i.e.  $\ln(p(\tau+1+Ry_\tau | \bar{M}_{R,F,\tau}))$ , giving a high score to a density that assigns a high probability to the outcome that actually occurs. In the final assessment, the average logarithmic score calculated over the evaluation period therefore reflects the model's performance in terms of its ability to reflect the range of possible outcomes for the variables of interest (see Mitchell and Wallis (2011) for further discussion). Naturally, if the final assessment of the forecast performance of the average models is based on the logarithmic score, the weights used in the average model should also be based on the average logarithmic scores obtained over the training period for the individual models.<sup>6</sup>

**Economic forecast evaluation criteria** The MSE and logarithmic score measures represent purely statistical criteria that reflect the economic worth of forecasts only in particular circumstances. Recent years have seen a growing interest in a decision-based approach to the evaluation and comparison of forecasts. Here, forecast accuracy is judged according to its economic value given an explicitly defined decision-making context<sup>7</sup>. We believe that a judgement on the usefulness of real time data in forecasts of these variables should include an evaluation element that reflects the economic worth of the forecast. In what follows, we describe a fair bet exercise that provides an economic evaluation of forecast performance in forecasting the likely occurrence of a number of events involving (the logarithm of) output,  $y_t$ , and price level,  $p_t$ , which capture different features

---

<sup>6</sup>The definition of the average model changes according to the weights and evaluation criteria that will be subsequently employed and we might distinguish between  $\bar{M}_{R,F,\tau}^{MSE}$  and  $\bar{M}_{R,F,\tau}^{LS}$  depending on whether weights are based on MSE or logarithmic score, for example. As the weights used are usually clear from the context, we will suppress the superscript in what follows for notational convenience.

<sup>7</sup>This recognises that the statistical criteria used to evaluate models, typically measured using mean squared forecasting error (MSE), provide information on the economic value of their forecasts only under certain conditions. See Granger and Pesaran (2000) for an overview of this discussion.

of the output and price nowcasts; namely:

- *DROP1* :  $\{ (\tau+1+Ry_\tau - \tau+Ry_{\tau-1} < 0) \}$ ; i.e. a nowcast of one period of negative growth at  $\tau$  (calculated using post-revision measures);
- *DROP2* :  $\{ (\tau+1+Ry_\tau - \tau+Ry_{\tau-1} < 0) \cap (\tau+Ry_{\tau-1} - \tau+R-1y_{\tau-2} < 0) \}$ ; i.e. a nowcast of two consecutive periods of negative growth at  $\tau$  (calculated using post-revision measures);
- *BPEAK* :  $\{ \tau+1+Ry_\tau < \max(\tau+Ry_{\tau-1}, \tau+R-1y_{\tau-2}, \tau+R-2y_{\tau-3}, \dots) \}$ ; i.e. period  $\tau$  output lies below its previous peak level;
- *BTREND* :  $\{ \tau+1+Ry_\tau < \tilde{y}_\tau, \text{ where } \tilde{y}_\tau = \frac{1}{5}(\tau+R-1y_{\tau-2} + \tau+Ry_{\tau-1} + \tau+1+Ry_\tau + \tau+2+Ry_{\tau+1} + \tau+3+Ry_{\tau+2}) \}$ ; i.e. output lies below trend, defined as the centred five-period moving average of output;
- *TURN* :  $\{ \text{turning point observed at } \tau \}$ ; where, following Harding and Pagan (2005), a peak in output at time  $\tau$  is nowcast to occur when

$$\begin{aligned} (\tau+1+Ry_\tau - \tau+Ry_{\tau-1}) &> 0; & (\tau+1+Ry_\tau - \tau+Ry_{\tau-1}) - (\tau+Ry_{\tau-1} - \tau+R-1y_{\tau-2}) &> 0 \\ (\tau+2+Ry_{\tau+1} - \tau+1+Ry_\tau) &< 0; & (\tau+3+Ry_{\tau+2} - \tau+2+Ry_{\tau+1}) - (\tau+2+Ry_{\tau+1} - \tau+1+Ry_\tau) &< 0 \end{aligned}$$

and a corresponding definition holds for a trough;

- *IRRISE* :  $\{ (1.5\tau+1\pi_\tau + 0.5\tau+1+R\tilde{y}_\tau) > (1.5\tau\pi_{\tau-1} + 0.5\tau+R\tilde{y}_{\tau-1}), \text{ where } \tau+1+R\tilde{y}_\tau = \frac{1}{5}(\tau+R-1y_{\tau-2} + \tau+Ry_{\tau-1} + \tau+1+Ry_\tau + \tau+2+Ry_{\tau+1} + \tau+3+Ry_{\tau+2}) \text{ and } \tau+R\tilde{y}_\tau = \frac{1}{5}(\tau+R-2y_{\tau-3} + \tau+R-1y_{\tau-2} + \tau+Ry_{\tau-1} + \tau+1+Ry_\tau + \tau+2+Ry_{\tau+1}) \}$ ; i.e. a nowcast of an interest rate rise based on simple Taylor rule involving the nowcasts of inflation and the output gap.

The probability of these events occurring at  $\tau$  can be nowcast using the average models  $\overline{M}_{R,F,\tau}$ ,  $\overline{M}_{0,F,\tau}$ ,  $\overline{M}_{R,-1,\tau}$  and  $\overline{M}_{0,-1,\tau}$  and comparison of the probability forecasts again provides an indication of the usefulness of real-time data. A straightforward statistical evaluation of a model's event probability forecast,  $\pi_\tau$ , is through a contingency table approach. Here, the forecast probability is converted to a prediction on whether the event will happen or not ( $r_t = 1$  or 0 respectively) depending on whether the probability is greater or less than 0.5. The performance of the model can then

be described by the ‘hit rate’ (i.e. the proportion of accurate predictions) or the Kuipers Score (a statistic that takes values between  $-1$  and  $1$  and summarises the degree of correspondence between predictions and outcomes similar to a correlation coefficient). Both statistics are calculated based on table 5.1 showing a standard contingency matrix of realisations of events and model forecasts,

Table 5.1: Contingency Matrix		
	Event Occurs	
Event Forecast	Yes;	No;
Yes; $r_t = 1$	$T_{by}$	$T_{gy}$
No; $r_t = 0$	$T_{bn}$	$T_{gn}$

where  $T_{by}$  represents the number of times the model correctly predicted the occurrence of the event,  $T_{bn}$  represents the number of times the model incorrectly predicted that the event will not occur,  $T_{gy}$  represents the number of times the model incorrectly predicted that the event will take place when in fact the event did not happen and  $T_{gn}$  gives the number of times the model correctly predicted that the event will not take place. The Hit rate is given by:

$$HIT = \frac{T_{by} + T_{gn}}{T_{by} + T_{gy} + T_{bn} + T_{gn}}. \quad (5.10)$$

The Kuipers score statistic is given by:

$$KS = H - F, \quad (5.11)$$

where  $H$  is the fraction of events that have taken place and the model correctly predicted their occurrence while  $F$  is the fraction of events that have not taken place and had been incorrectly forecast to have occurred (the "false alarm" rate, see Granger and Pesaran (2000)). In terms of the contingency table,  $F = \frac{T_{gy}}{T_{gy} + T_{gn}}$  and  $H = \frac{T_{by}}{T_{by} + T_{bn}}$ .

In addition, formal tests can be applied against the null that there is no relationship between the outcome and the predictions. Pesaran and Timmermann (1992) provide a test of a model’s ability to predict the occurrence of an event based on the proportion of occasions in which the model correctly predicts that the event will or will not happen. The test assumes the outcomes are uncorrelated over time, however, and Pesaran and Timmermann (2010) propose an alternative approach based on canonical regression methods to test for predictability in the presence of serial

dependence. Both approaches are purely statistical in the sense that they provide a test of the null of no predictability and take no account of the potential benefits of accurately predicting the event. They evaluate the model from the perspective of the producer of the forecast.

**Assigning economic significance to events** A more ‘economic’ evaluation might be based on an explicit investment scenario in which a model is used to predict the outcome of one of the events listed above. This attempts to evaluate the model from the perspective of the user of the model’s forecasts. One example of such an investment scenario is where an investor pays a fixed charge each period to participate in a scheme, predicts whether the event will occur or not and receives a payment if the prediction turns out to be true. Alternatively, the scenario might be asymmetric in that a payment is received only if the investor correctly predicts that the event *will* occur. In either scenario, the profits obtained from decision-making directly measure the economic value of the model over the evaluation period.

To formalise the ideas, note that any event defined at  $\tau$  as a set of outcomes involving outputs and/or prices up to  $h$  periods ahead can be written as  $R(\mathbf{X}_{\tau+1,\tau+h})$ . The probability that the event occurs is

$$\text{probability of event} = \pi_\tau = \int_R \Pr(\mathbf{X}_{\tau+1,\tau+h} | \mathbf{X}_{1,\tau}, \bar{M}_{R,F,\tau}) d\mathbf{X}_{\tau+1,\tau+h}. \quad (5.12)$$

In a simulation exercise, the forecast probability is obtained simply as the proportion of the simulations in which the event is observed to occur. In a decision-making context, where an individual’s objective function  $\nu(r_\tau, R(\mathbf{X}_{\tau+1,\tau+h}))$  depends on the outcome of a choice variable  $r_\tau$  and the occurrence of the recessionary event, the decision-maker’s problem can be written as

$$\max_{r_\tau} \left\{ \int \nu(r_\tau, R(\mathbf{X}_{\tau+1,\tau+h})) \Pr(\mathbf{X}_{\tau+1,\tau+h} | \mathbf{X}_{1,\tau}, \bar{M}_{R,F,\tau}) d\mathbf{X}_{\tau+1,\tau+h} \right\}. \quad (5.13)$$

In terms of the simulations, the decision involves simply choosing the value of  $r_\tau$  that maximises the value of the objective function when averaging across the simulations. We can denote the optimal value of the choice variable chosen using model  $\bar{M}_{R,F,\tau}$  by  $r_{R,F,\tau}$ . Pesaran and Skouras (2002) then suggest using the statistic

$$\bar{\Psi}_{R,F,\tau} = \sum_{\tau=T}^{T+k} \nu(r_{R,F,\tau}, R(\mathbf{X}_{\tau+1,\tau+h})), \quad (5.14)$$

calculated over an out-of-sample evaluation period  $T, \dots, T+k$  and based around the values of  $r_{R,F,\tau}$  chosen using model  $\overline{M}_{R,F,\tau}$  in each period. Similar statistics can be calculated for any other model, with associated optimal choice variable, and these provide the basis of a comparison of the forecast performance of the models on economic grounds.

The payout contingencies relating to the symmetric and asymmetric bets described above are summarised as:

Payout contingencies for outcomes of a symmetric fair bet			Payout contingencies for outcomes of an asymmetric fair bet		
	Recession Occurs			Recession Occurs	
Recession Forecast	<i>Yes</i>	<i>No</i>	Recession Forecast	<i>Yes</i>	<i>No</i>
<i>Yes</i>	$s - 1$	$-1$	<i>Yes</i>	$s - 1$	$-1$
<i>No</i>	$-1$	$s - 1$	<i>No</i>	$0$	$0$

where the fixed charge is 1 unit (payable every period in the symmetric case and only when participating in the asymmetric case) and where  $s$  represents the payment received with a successful prediction. In other words, in the symmetric case, it is assumed that the investor gambles £1 predicting whether the event will take place or not. If the prediction is correct, the investor gets a payout of £ $s$  but otherwise, loses the £1. In the asymmetric version, the investor only gambles with a £1 stake if she believes that the event will take place and wins £ $s$  if the prediction is correct. The bet is fair in each case because the payout is chosen so that the investor would break even if she made the choice randomly on the basis of the unconditional probability,  $p_u$ , that the event occurs. In the symmetric investment scenario this requires that expected profit  $(s-1)(p_u^2 + (1-p_u)^2) - 2p_u(1-p_u) = 0$ , i.e.  $s = \frac{1}{2p_u^2 - 2p_u + 1}$ .<sup>8</sup> Alternatively, it requires  $(s-1)p_u^2 - p_u(1-p_u) = 0$ ; i.e.  $s = \frac{1}{p_u}$  in the asymmetric case investment scenario. If the model's forecast probability is  $\pi$  and if the investor bets on recession when  $\pi$  exceeds some critical value  $\pi^c$ , then the decision to bet on recession or not

---

<sup>8</sup>If 'YY' indicates a recession is forecast and it occurs, 'YN' indicates a recession is forecast but it is not realised, and so on, then, the probability of 'YY' and 'NN' is equal to  $p_u^2$  and  $(1-p_u)^2$  respectively while the probability of 'YN' and 'NY' is equal to  $p_u(1-p_u)$ .

$(r_{R,F,\tau} = 1 \text{ or } 0)$  is equivalent to choosing the critical value. In the symmetric case, the investor's expected end-of-forecast-period wealth corresponding to  $\nu(r_\tau, R(\mathbf{X}_{\tau+1,\tau+h}))$  in (5.13) is given by:

$$E[W_{\tau+h}] = \begin{cases} (s-1)\pi - (1-\pi) & = \frac{\pi}{2p_u^2-2p_u+1} - 1 & \text{if } \pi > \pi^c \Leftrightarrow r_{R,F,\tau} = 1 \\ (s-1)(1-\pi) - \pi & = \frac{1-\pi}{2p_u^2-2p_u+1} - 1 & \text{if } \pi < \pi^c \Leftrightarrow r_{R,F,\tau} = 0 \end{cases}$$

and maximum expected wealth is achieved by choosing a critical value of  $\pi^c = 0.5$  since  $\frac{\pi}{2p_u^2-2p_u+1} > \frac{1-\pi}{2p_u^2-2p_u+1}$  if  $\pi > 0.5$  and vice versa if  $\pi < 0.5$ . In the asymmetric case, wealth is given by

$$E[W_{\tau+h}] = \begin{cases} (s-1)\pi - (1-\pi) = \frac{\pi}{p_u} - 1 & \text{if } \pi > \pi^c \Leftrightarrow r_{R,F,\tau} = 1 \\ 0 & \text{if } \pi < \pi^c \Leftrightarrow r_{R,F,\tau} = 0 \end{cases}$$

and maximum expected wealth is achieved by choosing a critical value of  $\pi^c = p_u$  since  $\frac{\pi}{p_u} - 1 > 0$  if  $\pi > p_u$ . In both cases, model  $\overline{M}_{R,F,\tau}$  can be used to predict the occurrence of a recession or not in each observation through the evaluation period and, depending on the actual outcome, this will generate a sequence of financial returns that can again be used to judge the model as in (5.14). Carrying out the same exercise for models  $\overline{M}_{R,-1,\tau}$ ,  $\overline{M}_{0,F,\tau}$  and  $\overline{M}_{0,-1,\tau}$  provides a \$ value for each model which are comparable across models and which conveys directly the economic usefulness of each of these models (and of the different elements of the real-time data).

The criterion function clearly depends on the decision-making context. However, the use of a range of alternative investment scenarios can provide a sense of the robustness of the model evaluation to changes in the detail of the scenario. The break-even payment is largest when  $p_u = 0.5$  in the symmetric case reflecting the conservative economic setting in which returns are greatest when correctly predicting an event that is as likely to occur as not. In the asymmetric scenario, however, the payment becomes infinitely large as  $p_u \rightarrow 0$  so that there are large gains to be made from predicting unlikely events in this case. These two scenarios provide quite different economic contexts therefore. A model that performs well in both would have established itself as useful in quite different situations while good performance in one but not another provides information on the contexts in which the model is and is not useful.

### 5.3 Nowcasting Recessionary Events using US Real-Time Data

The empirical work of the chapter considers nowcasts of output and price outcomes and business cycle features based on the first-release and revised measures of output and price and the direct measures of output and price expectations obtained from the real-time datasets of the Federal Reserve Bank of Philadelphia available at [www.phil.frb.org/econ/forecast/](http://www.phil.frb.org/econ/forecast/). The officially-released ‘backward-looking’ series consist of 196 quarterly vintages of data; the first was released in 1965q4 and the final vintage used in this chapter is dated 2014q3. All vintages include observations dated back to 1947q1<sup>9</sup>. The ‘forward-looking’ series are the experts’ forecasts on output provided in the *Survey of Professional Forecasters* (SPF) from 1968q4–2014q3. The forecasts in the SPF are made around the mid-point of quarter  $t$  and include nowcasts of the current quarter and forecasts of up to four quarters ahead. In fact, the data on US real GDP in quarter  $t - 1$  are released for the first time at the end of the first month of quarter  $t$  so the first-release information on the previous quarter’s output is available to the professional forecasters at the time they make their forecasts. Nevertheless, it is reasonable to assume  ${}_ty_{t-1}$  and  ${}_ty_{t+f}^e$ ,  $f = 0, \dots, 4$  are determined simultaneously when working at the quarterly frequency.

Assuming that output data is revised for three quarters while the price series is not revised, then actual quarter-on-quarter output growth at time  $t$ , can be measured by the post-revision series  ${}_{t+4}y_t - {}_{t+4}y_{t-1}$ . Figure A1 in the appendix illustrates the nature of the output series under investigation. The output series has an average annualised rate of 0.61% (with standard deviation of 0.83%) and is plotted in Figure A1 alongside the first-release and first-revision series. Figures A2-A4 plot the revisions of output series showing that revisions are small on average but the first and second revisions have a range of  $[-1.58\%, 1.63\%]$  and  $[-1.23\%, 1.55\%]$ , and standard deviation of 0.41% and 0.34%, respectively and so are often of a similar order of magnitude to the actual growth figures themselves. Figures A2-A4 also demonstrate that there are occasionally some very

---

<sup>9</sup>The analysis distinguishes between standard ‘revisions’ and once-and-for-all ‘benchmark adjustments’ arising out of the re-definition or reclassification of a series. The latter are announced in advance and we assume these are entirely taken into account in forecasting and decision-making. To do this, we adjust the data by splicing the pre- and post-benchmark-adjustment series to eliminate the effects prior to the analysis. Benchmark adjustments took place in 1976q1, 1981q1, 1986q1, 1992q1, 1996q1, 1999q4, 2004q1 and 2009q3.

large revisions, with a relatively large number in excess of 0.5% occurring during the late seventies and mid-eighties and a large number less than  $-0.5\%$  in the early eighties and after 2007. The fact that these episodes coincide with periods of unusually strong or weak growth suggests that the measurement errors are (understandably) related to business cycle conditions and suggests that revisions may be more or less useful in forecasting growth outcomes at different times.

The expectations series obtained from the SPF are shown in figure A5, again set against actual post-revision growth. This figure shows that the output expectations series also display some volatility but they move more conservatively than the actual growth series itself. The conservatism becomes more pronounced as the forecast horizon grows so that four-period-ahead survey expectations rarely move outside the  $[0.5\%, 1.0\%]$  range, especially over the latter half of the sample. Defining expectational errors observed in the SPF series by  ${}_{t+4}y_t - {}_{t-f}y_t^e$  for  $f = 0, \dots, 3$ , i.e. comparing the post-revision series to the survey forecasts for up to 3 quarters earlier,<sup>10</sup> figure A6 plots the expectational errors directly, showing some very large errors in the four-period-ahead forecasts.

As far as the price series is concerned, we assume that it is not revised (indeed revisions are minimal) so actual inflation is measured by the growth in GDP deflator based on the most-recent vintage i.e.  ${}_{TP}p_t - {}_{TP}p_{t-1}$ . Figure A7 shows the irregularities the inflation series has experienced with the dramatic increase in inflation during the turmoil years of the 1970s and fast disinflation during the 1980s. The price series has an average annualised rate of 0.91% (with standard deviation of 0.64%). As figure A7 conveys, the revisions are small as the first-release inflation series and inflation based on the most-recent vintage are almost identical.

The expectations of the price series as given by SPF are shown in figure A8. As with the output series, while the prices expectations exhibit some volatility, they fluctuate more moderately than actual inflation series. Indeed, the four step ahead expected inflation displays less pronounced fluctuations. For example, during the early 1970s when actual inflation exceeded 3.25%, the four step ahead expected inflation remained well below 1.5% and followed actual inflation's upward movements in a much more conservative rate. Defining expectational errors observed in the SPF series by  ${}_{t+4}p_t - {}_{t-f}p_t^e$  for  $f = 0, \dots, 4$ , figure A9 shows great variability in errors. The figure shows

---

<sup>10</sup>This is the appropriate measure of 'expectational error' only if the survey participants report predictions of actual, post-revision output in their returns, not the predictions of the first-release measure.



large periods when the expectational errors are close to zero while substantial spikes are observed especially for errors in the four-period-ahead price forecasts.

### 5.3.1 Real-time evaluation of point and density forecasts

The purpose of the empirical work is to find whether the information contained in the revision and survey data is useful in nowcasting the actual output series and/or the business cycle features reflected in these series. All of the models that we estimate can be accommodated by the average model  $\overline{M}_{R,F,\tau}$  defined in (5.8), with  $r = 0, 1, 2, 3$  and  $f = -1, 0, 1, 2, 3$ , and in (5.2)-(5.4)<sup>11</sup>. Hence, twenty versions of the model in (5.2)-(5.4) are estimated with the most general including three revisions and survey forecasts up to three quarters ahead in addition to the first release data, while the most simple version of the model is the ‘conventional model’ which uses the first-release data only.

This exercise begins by estimating the average model  $\overline{M}_{R,F,1991q2}$  defined in (5.8) and using the real time data available for 1968q4 – 1991q2 using the 80-quarter period 1968q4 – 1988q3 in estimation and holding back the 12 quarters’ data for 1988q3 – 1991q2. Each of the twenty underlying models are used to produce forecasts of the various measures of output, including for example, the one-step ahead forecast of the first release measure of contemporaneous output, 1988q4y1988q3, say, and the four-period-ahead forecast of the post-revision measure of contemporaneous output, 1989q3y1988q3. For the purpose of obtaining the model weights, we focus here on the forecast of post-revision measure, comparing this to the post-revision outcome observed during the training period. The twenty models are then estimated over the 81-quarter period 1968q4 – 1988q4 and the forecast of the post-revision measure 1989q4y1988q4 is obtained and compared to the observed outcome. This is repeated over the whole training period, moving recursively and judging the relative performance of the twenty models each time to obtain the set of weights,  $w_{r,f,1991q2}$ , defined in (5.6) for the RMSE and (5.7) for the log scores, for  $r = 0, \dots, 3$  and  $f = -1, \dots, 3$ . Moving on one period, this entire exercise can then be repeated over the sample 1969q1 – 1991q3 using 1988q4 – 1991q3 as the training period, to derive the set of weights  $w_{r,f,1991q3}$  and so on to the final vintage data

---

<sup>11</sup>In other words, for the empirical exercise,  $R = 3$  and  $F = 3$ .



Figure 5-2: The real-time assessment of the usefulness of expectations survey data in post revision output forecasts, based on the  $\mu_{\tau}^r$  statistic.

2013q2.

Figure 5.2 shows  $\mu_{\tau}^r$  defined in (5.9) i.e. the weighted average of the revision length of the models included in the model  $\overline{M}_{R,F,\tau}$  for  $\tau = 1991q3, \dots, 2014q3$  based on the models' forecast and judged according to their RMSE (when using weights as in (5.6)) and to logarithmic scores (when using weights as in (5.7)). The plot based on RMSE shows a greater degree of stability: the average revision horizon is around 1.5 and lies in the range  $[1.00, 2.00]$  for *nearly* all the sample. Although the statistic exhibits some variation through time, it also reflects the finding that, when using RMSE as the criterion, many models appear to perform equally well so that the average of their revision lengths is mid-way between zero and three, the minimum and maximum values. In contrast, the plot based on log score weights is much more discerning, showing a very low average revision length -i.e. with few revisions used - during the first part of the evaluation period (weighted average revision length remained below 1.5 for the first part of the sample) but rising above 1.5

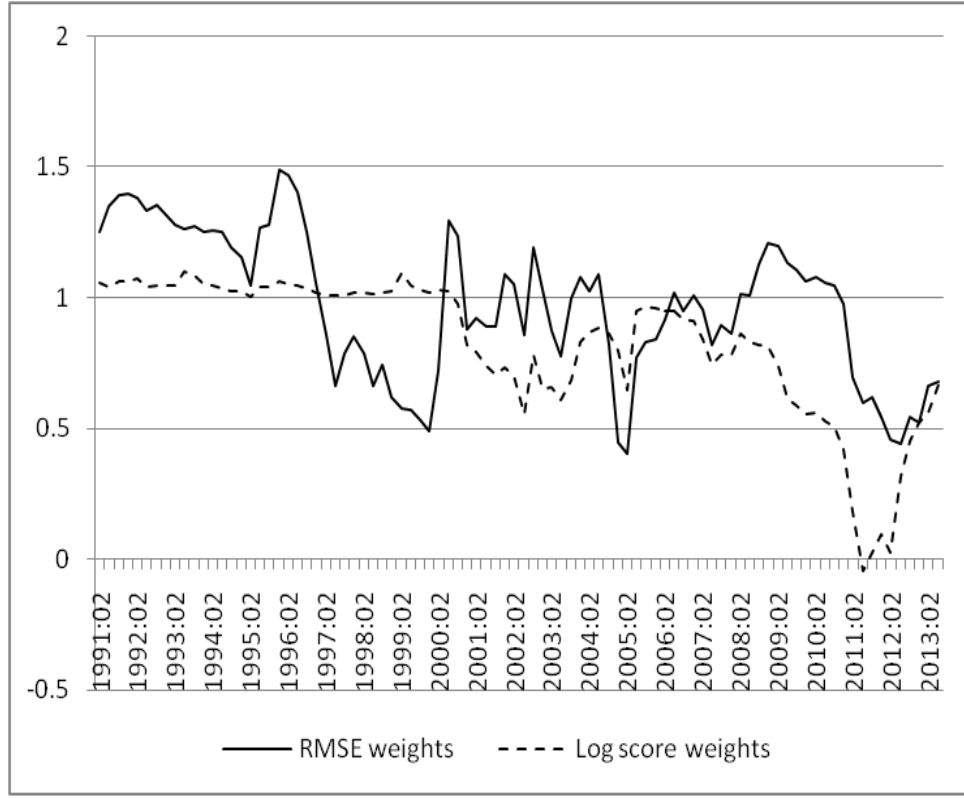


Figure 5-3: The real-time assessment of the usefulness of expectations survey data in post revision output forecasts, based on the  $\mu_{\tau}^f$  statistic.

during the early 2000's and to close to 3.0 - making full use of revisions - over the evaluation period after 2007. As noted above, there were a number of large revisions in the output data released in the early 2000's and again in the years following the financial crisis and it appears that the average model adjusts to exploit the extra information contained in the revisions at this time, placing more weight on models that include the long revisions.

Figure 5.3 shows  $\mu_{\tau}^f$  defined in (5.9) i.e. the weighted average of the survey horizon of the models included in the model  $\overline{M}_{3,3,\tau}$  for  $\tau = 1991q1, \dots, 2014q3$  based on the models' forecast and judged according to their RMSE (i.e. using weights as in (5.6)) and to logarithmic scores (i.e. using weights as in (5.7)). The main message from figure 5.3 is that survey data are relatively important over the whole sample and the statistic never becomes negative, indicating that at least contemporaneous measures of output and price level are useful in nowcasting and forecasting. The weights based on RMSE fluctuate around 1, mid-way between the minimum and maximum values

of  $-1$  and  $3$ , once more reflecting the difficulty in discriminating between models according to their point forecasts. Nevertheless, at the end of the evaluation period, the weighted average forecast horizon based on RMSE weights declines substantially, showing that agents used information from surveys only for the contemporaneous values of output and price outcomes. The weights based on log scores show greater variability, taking values which are close to zero towards the end of the sample. This pattern is less easy to interpret, although given the timing, we might speculate that the surveys could become less reliable in density forecasting during recession or that the relative conservatism in survey data may force a spurious precision in the density forecasts during periods of volatility. In any case, the real time evaluation exercise indicates that survey data that provide information on output and price expectations for the current and one-step ahead values can be useful for forecasting but their usefulness changes over time.

Figures 5.4 and 5.5 provide some further insight on this shift in the weights over time, showing the observed post-revision output series alongside the point forecasts and 5th/95th percentile of the forecast densities for the most general model  $\overline{M}_{R,F,\tau}$  and for the simplest model  $\overline{M}_{0,-1,\tau}$ , based on logarithmic score weights, during two illustrative episodes. Figure 5.4, which relates to the period 1994q1 – 1995q4 at the beginning of the evaluation period, shows that the point forecasts of the two models are broadly the same. However, the forecasts density is rather narrower for the  $\overline{M}_{R,F,\tau}$  model, suggesting that models that do not make use of revisions and survey data embed higher uncertainty about the true value of post-revision outcome, so that the  $\overline{M}_{R,F,\tau}$  model outperforms the simpler  $\overline{M}_{0,-1,\tau}$  model in terms of log score. Figure 5.5 shows that over the period 2009q2 – 2011q1, when there were some large revisions in the data, the point forecasts of the  $\overline{M}_{R,F,\tau}$  model are closer to the actual than those from model  $\overline{M}_{0,-1,\tau}$  and the densities are also wider so that the observed outcome lies in the body of the forecast density much more often. This illustrates the idea that, by placing more weight on the models including long revisions, the average model adjusts to incorporate the information contained in the revisions during the periods when revisions become significant.

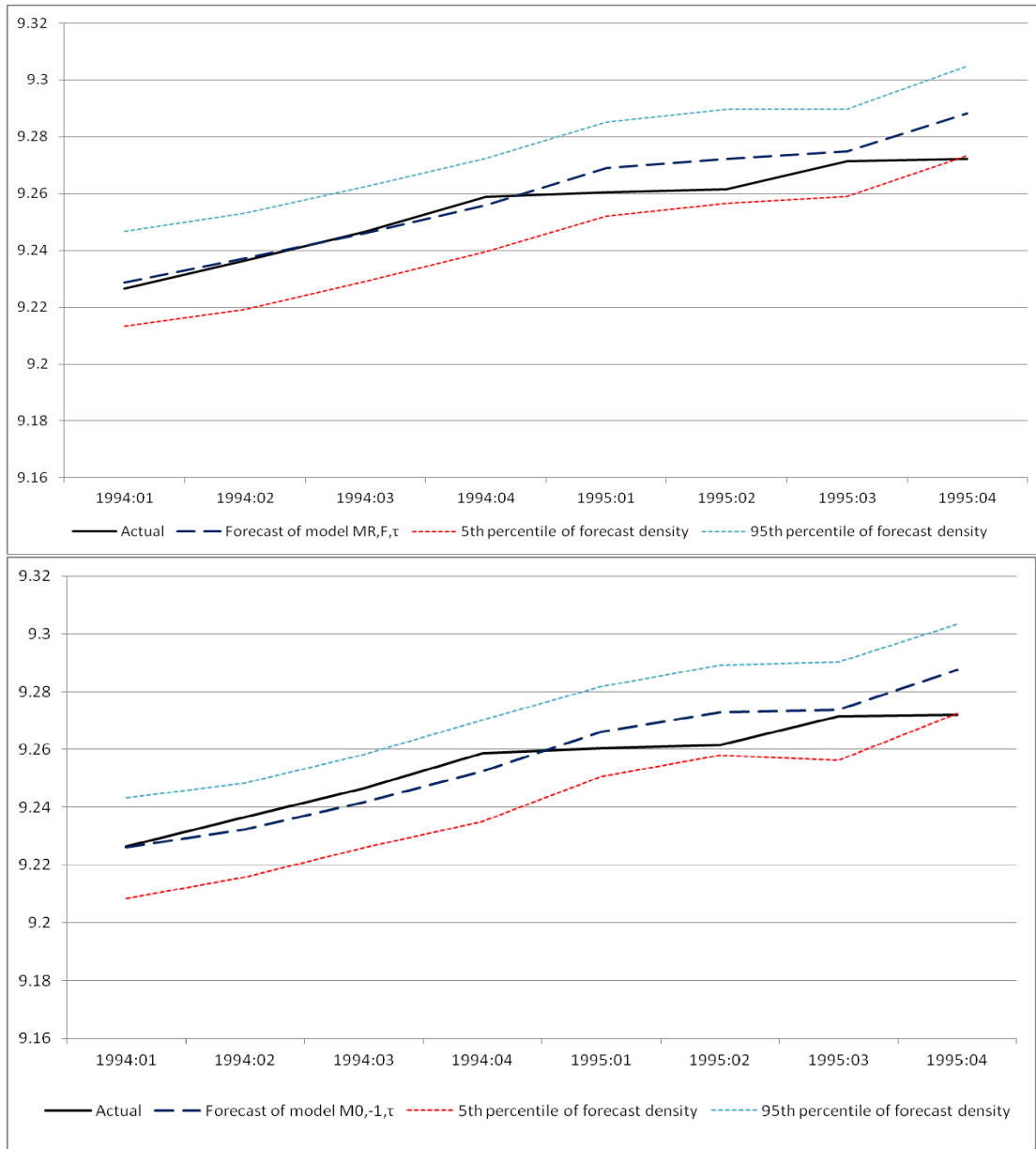


Figure 5-4: Mean, 5th percentile and 95th percentile of forecast density of models  $\bar{M}_{R,F,\tau}$  and  $\bar{M}_{0,-1,\tau}$  under the logarithmic score weight scheme over the period 1994q1 – 1995q4 (Post-revision outcome).

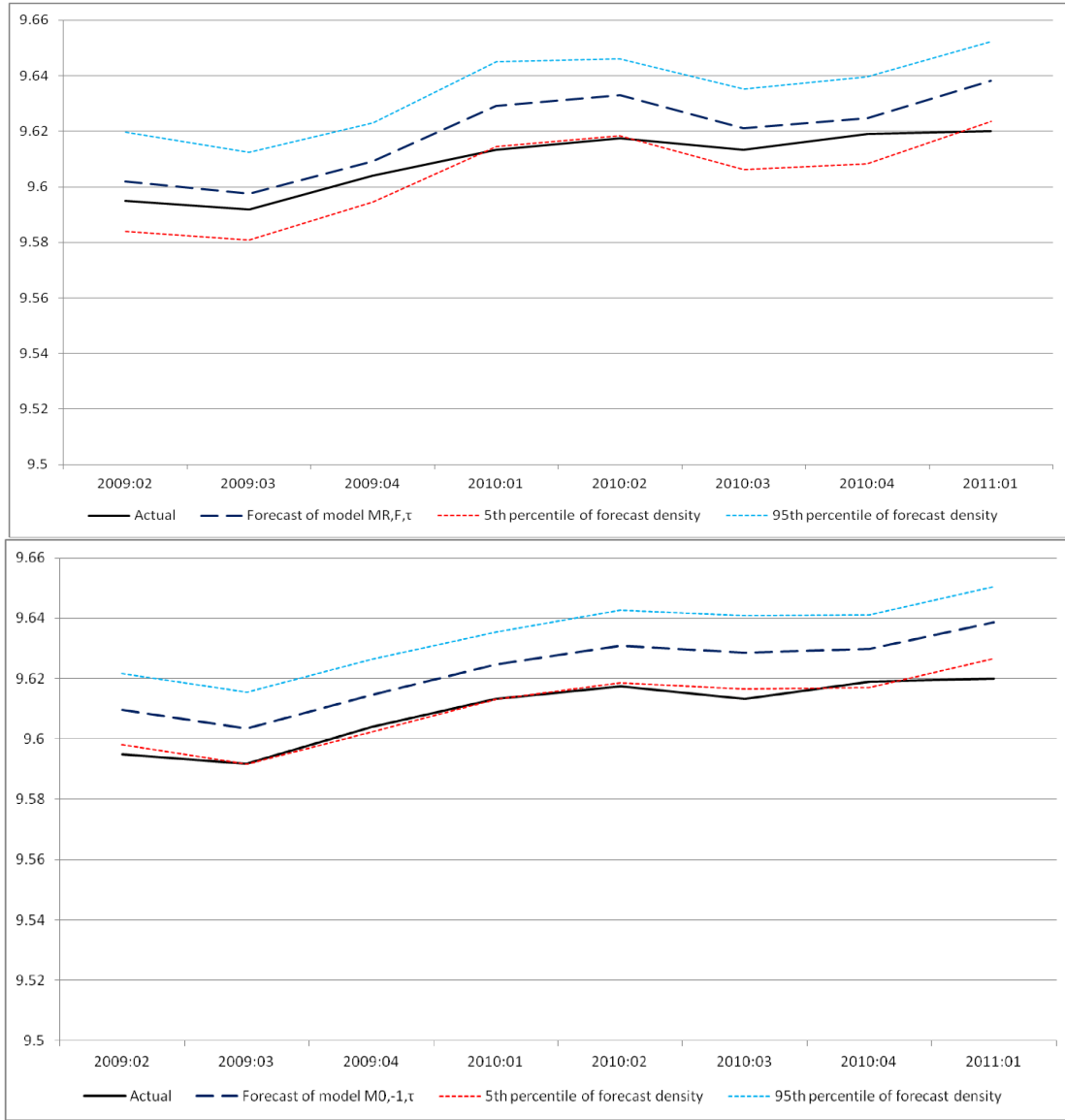


Figure 5-5: Mean, 5th percentile and 95th percentile of forecast density of models  $\bar{M}_{R,F,\tau}$  and  $\bar{M}_{0,-1,\tau}$  under the logarithmic score weight scheme over the period 2009q2 – 2011q1 (Post-revision outcome).

### 5.3.2 Final-assessment of model predictability based on economic events

The shifting weights over time provide insights on the usefulness of revisions and survey data in forecasting as would be judged at the time. The ‘final assessment’ statistics of Table 5.1 judge their usefulness over the whole evaluation period by comparing the forecast performance of four alternative average models which are more or less constrained in their use of the revisions and survey data. Specifically here, we compare the performance of: (i) the general average model discussed above,  $\overline{M}_{R,F,\tau}$  for  $\tau = 1991q2 - 2013q2$  which uses the revisions and survey data as the estimated weights indicate; (ii) the average model  $\overline{M}_{R,-1,\tau}$ , obtained from models that differ in their use of revisions but make no use of the survey data at all; (iii)  $\overline{M}_{0,F,\tau}$  making no use of revisions; and (iv) the ‘conventional’ meta model,  $\overline{M}_{0,-1,\tau}$ , making no use of revisions data or survey data. In principle, we could conduct a separate forecast evaluation at every forecast horizon and for each of our output measures (i.e. the first-release measure and various revisions and survey expectation measures at different future dates). In what follows, we focus on the four-period-ahead forecast of the post-revision measure of contemporaneous output which is a natural way of thinking of the ‘nowcast of current actual output’.

The results of table 5.2 show the ‘conventional’ average model,  $\overline{M}_{0,-1,\tau}$  has an average RMSE of 1.14% when judged over the whole evaluation period 1991q2 – 2013q2. The three average models  $\overline{M}_{R,F,\tau}$ ,  $\overline{M}_{R,-1,\tau}$  and  $\overline{M}_{0,F,\tau}$  all outperform the conventional model, with smaller average RMSEs reported in each case but only  $\overline{M}_{R,-1,\tau}$  shows a statistically significant improvement according to the Giacomini-White (GW) test of equal forecasting performance (where the GW tests are performed using Newey-West robust standard errors with automatic selection for bandwidth). When weights are chosen according to the log scores, models  $\overline{M}_{R,F,\tau}$  and  $\overline{M}_{R,-1,\tau}$ , both of which make use of the revisions data, show an improvement in the log-score over that of the conventional average model, but only  $\overline{M}_{R,-1,\tau}$  shows a statistically significant improvement. Average model  $\overline{M}_{0,F,\tau}$ , which uses surveys but not revision data, actually has a deterioration in forecast performance compared to the conventional model, but the difference is not statistically significant. Hence, the ‘final evaluation’ results suggest it is a good idea to include real-time data when forecasting output although the argument for the use of revisions is more compelling than for the use of surveys. In every case, it

is a good idea to take into account the fact that the real-time data might be more or less useful at different times.

Table 5.2: RMSE and Average Logarithmic Predictive Scores for Output Growth Nowcasts:

1991q2 – 2013q2 (Post-Revision Outcome)

*Actual RMSE and Average Logarithmic scores for model  $\overline{M}_{0,-1,\tau}$*

*Scaled difference from model  $\overline{M}_{0,-1,\tau}$  for other models*

Average Model	RMSE	Log-score
	Actual for $\overline{M}_{0,-1,\tau}$	
$\overline{M}_{0,-1,\tau}$	0.0114	-2.562
	Relative to $\overline{M}_{0,-1,\tau}$	
$\overline{M}_{R,-1,\tau}$	-0.0010**	0.7041***
$\overline{M}_{0,F,\tau}$	-0.0010	-0.2640
$\overline{M}_{R,F,\tau}$	-0.0012*	0.7503

Note: The average model  $\overline{M}_{R,F,\tau}$  is as defined in (5.8). In particular, model  $\overline{M}_{R,F,\tau}$  denotes the unrestricted average model which includes use of the full available revisions and survey data on expectations for  $\tau = \underline{\tau}, \dots, T$ ;  $\overline{M}_{0,F,\tau}$  denotes the model where no use of the revisions data is made for  $\tau = \underline{\tau}, \dots, T$ ;  $\overline{M}_{R,-1,\tau}$  denotes the model where no use is made of the survey data for  $\tau = \underline{\tau}, \dots, T$  and  $\overline{M}_{0,-1,\tau}$  denotes the model where no use is made of either the revisions or survey data throughout (i.e. based on the conventional real time model), for  $\tau = \underline{\tau}, \dots, T$ . A ‘\*’ denotes significance at the 10% level, ‘\*\*’ denotes significance at 5% level and ‘\*\*\*’ significance at the 1% level of the Giacomini-White (2006) test of equal forecast performance testing whether the RMSE and the logarithmic predictive score are significantly smaller or larger, respectively, than the corresponding statistics from model  $\overline{M}_{0,-1,\tau}$ .

### Evaluation of event probabilities and fair bet outcomes

We have argued that models’ forecast performance might also be judged by their ability to predict recession and to enhance decision-making. We also consider the models against this criterion then, with predictions again based on density forecasts of the post-revision output measures and outcomes measured by the realised post-revision series. In what follows, we use six events, described in



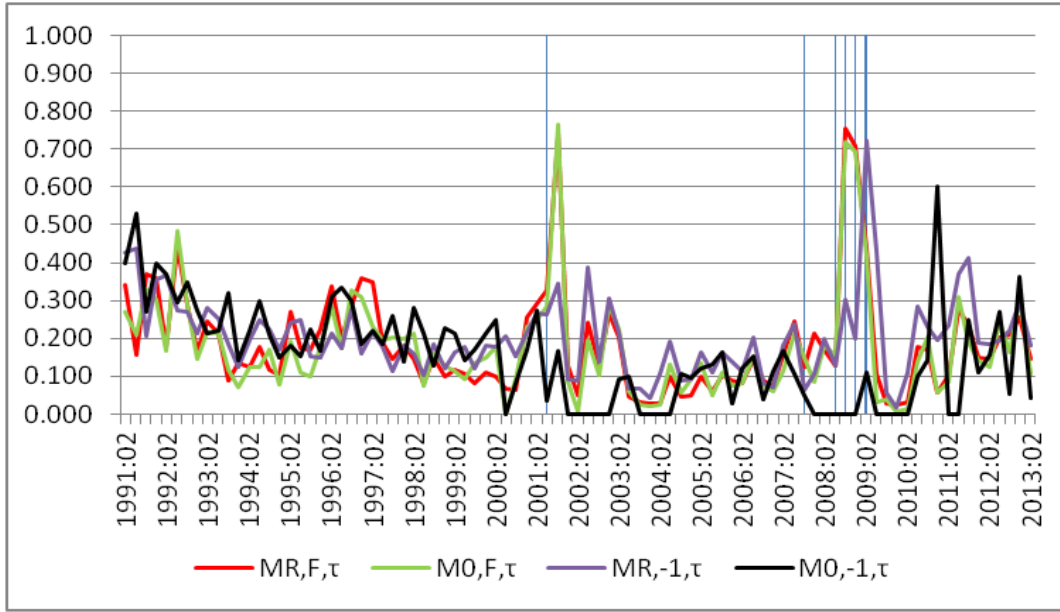


Figure 5-6: Probability forecasts of one period of negative output growth (post-revision outcome), as generated by average models  $\bar{M}_{0,-1,\tau}$ ,  $\bar{M}_{0,F,\tau}$ ,  $\bar{M}_{R,-1,\tau}$  and  $\bar{M}_{R,F,\tau}$ , under the RMSE weight scheme. Vertical lines denote when the event has taken place.

section 5.2.2, that capture different recessionary features of the business cycle at time  $\tau$ . Generally, and in the absence of specified pay-out contingencies, a forecasted probability of recession that exceeds 0.5 is interpreted as predicting that recession will occur. Figures 5.6 and 5.7 illustrate the type of results obtained, showing the forecast probabilities of one period of negative output growth (DROP1) according to each model under the two weight schemes, as well as showing when the event actually occurred. As is clear, this event occurs relatively infrequently with a run of consecutive events during the financial crisis.

What becomes immediately apparent from figure 5.6 is that, under the RMSE weight scheme, the quasi model  $\bar{M}_{0,-1,\tau}$  is exceptionally bad in nowcasting one period of negative output growth as it fails to identify the occurrence of the event in all cases that it takes place. This makes model  $\bar{M}_{0,-1,\tau}$  redundant for the nowcasting of one period of negative output growth, suggesting that useful information is excluded from the model. Even more importantly, model  $\bar{M}_{0,-1,\tau}$  often results in false alarms since it signifies that the event will take place at times when the event does not actually take place (such as in 1991q3 and 2011q1). While the event takes place six times during the

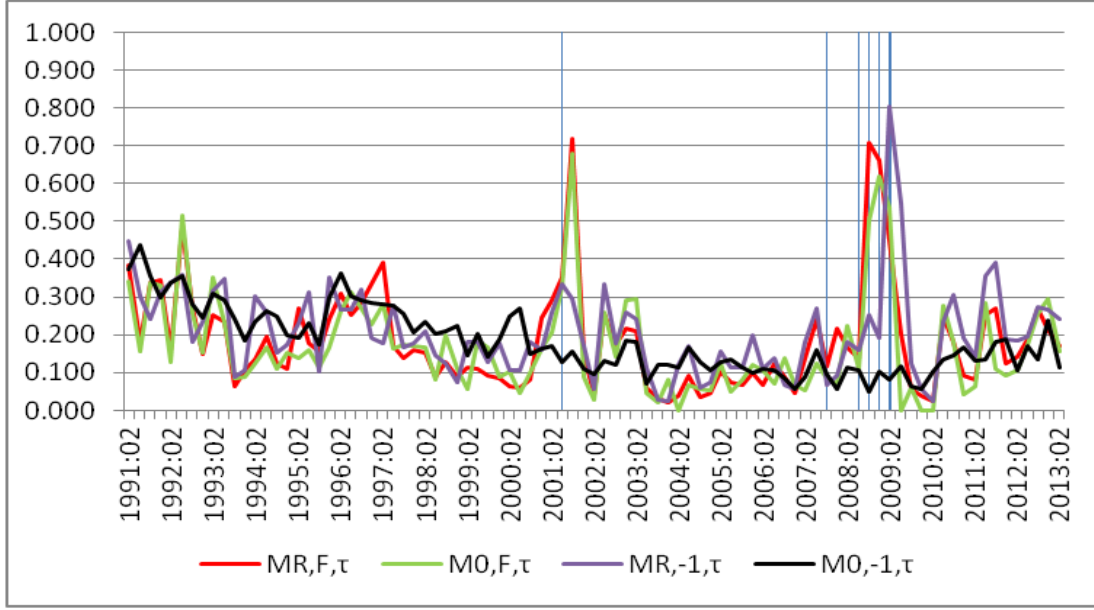


Figure 5-7: Probability forecasts of one period of negative output growth (post-revision outcome), as generated by average models  $\bar{M}_{0,-1,\tau}$ ,  $\bar{M}_{0,F,\tau}$ ,  $\bar{M}_{R,-1,\tau}$  and  $\bar{M}_{R,F,\tau}$ , under the logarithmic score weight scheme. Vertical lines denote when the event has taken place.

sample period, all average models under the RMSE weight scheme fail to identify the occurrence of the event in three cases (2001q3, 2007q4, 2008q3). Despite this insufficiency, average models that make use of survey data appear to predict the occurrence of events with greater confidence, delivering probability forecasts that are close to 1 in a number of cases where the event actually takes place. Moreover, models  $\bar{M}_{R,F,\tau}$  and  $\bar{M}_{0,F,\tau}$  attain low predicted probabilities, mostly below 0.4, when the event does not take place. Very similar conclusions are drawn from figure 5.7.

Tables 5.3a and 5.3b provide measures of the extent to which the models meet the challenges of forecasting recessions defined in the various ways. For Tables 5.2a and 5.2b, the 89 predictions and outcomes observed over 1991q2 – 2013q2 are arranged into a two-by-two contingency table. Table 5.3a shows the proportion of forecasts that are correct, as shown in (5.10) for each model and Table 5.3b reports the Kuipers scores, as calculated based on (5.11). Table 5.3b also reports, in parentheses, the results of two tests described in Pesaran and Timmermann (2009): a static  $\chi^2$  test of whether a model's forecast performance is any better than would have been achieved guessing randomly based only on the unconditional probability of the event  $p_u$ ; and a dynamic version in

which the random guess also takes account of the possibility that the event is known to occur in runs.

The events DROP1, DROP2 and TURN occur relatively infrequently and so the hit rates (proportion correct) of Table 5.3a - which treat correct predictions of non-occurrence in the same way as correct predictions of occurrence - are high across all models as would be expected. Nevertheless, it appears that the average models  $\overline{M}_{R,-1,\tau}$  and  $\overline{M}_{R,F,\tau}$ , for the RMSE weight scheme, and  $\overline{M}_{R,F,\tau}$ , for the log score weight scheme, achieve the highest hit rates in most cases while in many cases, the model  $\overline{M}_{0,-1,\tau}$  attains the lowest hit rates. The inadequacy of the conventional model  $\overline{M}_{0,-1,\tau}$  and the superiority of average model  $\overline{M}_{R,F,\tau}$  matches the plots of Figures 5.6 and 5.7: Average model  $\overline{M}_{R,F,\tau}$  generates low event probabilities during most of the evaluation period but indicates a high probability of recession during the relevant quarters of the financial crisis. The dominance of the average model model  $\overline{M}_{R,F,\tau}$  is confirmed, and exaggerated, in the Kuipers scores of Table 5.3b which focus more on the models' ability to correctly predict the rare events. The Kuipers score also highlight the superiority of the average model  $\overline{M}_{0,F,\tau}$  under both weights schemes, validating the idea that survey data enhance model's ability to predict rare events. In general, the  $\chi^2$  tests also confirm that the performance of  $\overline{M}_{R,F,\tau}$  and  $\overline{M}_{0,F,\tau}$  is significantly better than would be achieved by chance, unlike the 'conventional' average model,  $\overline{M}_{0,-1,\tau}$ . The over-riding conclusion then is that the models provide a valuable tool for forecasting rare recessionary events and that the models which include the survey data typically perform best in predicting these rare events<sup>12</sup>.

---

<sup>12</sup>The results show that none of the models perform well in predicting event TURN. This is perhaps unsurprising given the complexity of the event. Nevertheless, this illustrates the important point that there are events that are difficult for any model to predict and that forecasters should consider when models are fit for purpose and when they are not.

Table 5.3a: Forecasting Events: Hit Rate: 1991q2 – 2013q2 (Post-Revision Outcome)									
	$p_u$	RMSE Weights				Log Score Weights			
		$\overline{M}_{0,-1,\tau}$	$\overline{M}_{R,-1,\tau}$	$\overline{M}_{0,F,\tau}$	$\overline{M}_{R,F,\tau}$	$\overline{M}_{0,-1,\tau}$	$\overline{M}_{R,-1,\tau}$	$\overline{M}_{0,F,\tau}$	$\overline{M}_{R,F,\tau}$
<i>DROP1</i>	7%	0.910	<b>0.943</b>	<b>0.943</b>	<b>0.943</b>	0.933	0.933	0.932	<b>0.943</b>
<i>DROP2</i>	5%	0.955	<b>0.966</b>	<b>0.966</b>	<b>0.966</b>	0.955	<b>0.966</b>	<b>0.966</b>	<b>0.966</b>
<i>BPEAK</i>	22%	0.899	<b>0.921</b>	0.910	0.899	<b>0.910</b>	0.876	0.887	0.887
<i>BTREND</i>	48%	0.550	0.505	0.550	<b>0.573</b>	0.595	0.550	<b>0.629</b>	0.550
<i>TURN</i>	2%	0.978	0.978	0.978	0.978	0.978	0.978	0.978	0.978
<i>IRRISE</i>	48%	0.606	0.674	<b>0.685</b>	0.629	0.584	0.640	0.662	<b>0.685</b>

Note: The average model  $\overline{M}_{R,F,\tau}$  is as defined in (5.8). Event DROP1 is one-period negative output growth; DROP2 is two successive periods of negative output growth; BPEAK is output level below previous peak; BTREND is output level below 5-period moving average; TURN is a turning point in output; IRRISE is interest rate rise by Taylor rule.  $p_u$  is the unconditional probability of the event for the period 1991q2 – 2013q2. Emboldened figures show the largest hit rate.

Table 5.3b: Forecasting Events: Kuipers Score: 1991q2 – 2013q2 (Post-Revision Outcome)									
	$p_u$	RMSE Weights				Log Score Weights			
		$\overline{M}_{0,-1,\tau}$	$\overline{M}_{R,-1,\tau}$	$\overline{M}_{0,F,\tau}$	$\overline{M}_{R,F,\tau}$	$\overline{M}_{0,-1,\tau}$	$\overline{M}_{R,-1,\tau}$	$\overline{M}_{0,F,\tau}$	$\overline{M}_{R,F,\tau}$
<i>DROP1</i>	7%	-0.024 (-, -)	0.166 (***, **)	<b>0.321</b> (***, *)	<b>0.321</b> (***, *)	0.000 (-, -)	0.154 (**, -)	0.309 (***, -)	<b>0.321</b> (***, *)
<i>DROP2</i>	5%	0.000 (-, -)	0.250 (***, **)	<b>0.488</b> (***, ***)	<b>0.488</b> (***, ***)	0.000 (-, -)	0.250 (***, **)	<b>0.488</b> (***, ***)	<b>0.488</b> (***, ***)
<i>BPEAK</i>	22%	0.550 (***, *)	0.650 (***, **)	<b>0.671</b> (***, *)	0.621 (***, -)	<b>0.600</b> (***, -)	0.521 (***, -)	0.571 (***, -)	0.571 (***, -)
<i>BTREND</i>	48%	0.091 (-, -)	0.013 (-, -)	0.102 (-, -)	<b>0.148</b> (-, -)	0.194 (*, *)	0.106 (-, -)	<b>0.258</b> (**, **)	0.103 (-, -)
<i>TURN</i>	2%	0.000 (-, -)	0.000 (-, -)	0.000 (-, -)	0.000 (-, -)	0.000 (-, -)	0.000 (-, -)	0.000 (-, -)	0.000 (-, -)
<i>IRRISE</i>	48%	0.212 (-, -)	0.345 (***, ***)	<b>0.368</b> (***, ***)	0.256 (-, -)	0.159 (-, -)	0.277 (**, **)	0.322 (***, ***)	<b>0.367</b> (***, ***)

Note: The average model  $\overline{M}_{R,F,\tau}$  is as defined in (5.8).  $p_u$  is the unconditional probability of the event for the period 1991q2 – 2013q2; emboldened figures show the largest Kuipers score; the figures in parentheses (a,b) below the Kuipers Scores show the outcome of the static and dynamic versions of the Pesaran and Timmermann (2009) tests of no additional predictive power beyond that of the unconditional probability; a ‘\* \* \*’ indicates significance at 1% level, ‘\*\*’ indicates significance at 5% level, ‘\*’ indicates significance at 10% level, and ‘-’ indicates no significance at 10% level.

Table 5.4a and 5.4b provide the results of evaluating forecasting performance in the more sophisticated ‘fair bet’ decision-making context, reporting the returns achieved by a forecaster using each of the average models in the symmetric and asymmetric scenarios, described in (5.13) and (5.14). The actual return is reported for model  $\overline{M}_{0,-1,\tau}$  and the improvement over  $\overline{M}_{0,-1,\tau}$  is reported for other models. Table 5.4a relates to the symmetric fair bet in which the investor bets every period, gains the same payout for correctly predicting the occurrence and non-occurrence of events and compares the forecast recession probability against the same 0.5 threshold used in Tables 5.3a and 5.3b. Given that the set up is similar to that underlying the Kuipers score, it is not surprising that the results are very similar to those in Table 5.3b: the rank ordering of the models

obtained for each event is broadly the same with average models  $\overline{M}_{0,F,\tau}$  and  $\overline{M}_{R,F,\tau}$  dominating under both weight schemes. The results are a little different in Table 5.4b, where the asymmetric setup delivers a higher payout on events that are more rare and the forecast probability is compared against the unconditional probability. Under the RMSE weight scheme, the average model  $\overline{M}_{R,F,\tau}$  achieves the highest returns in most cases while under log score weights, average model  $\overline{M}_{0,F,\tau}$  is the dominant model and attains the highest returns in four out of six events considered. Again, the over-riding conclusion is that models which include survey data perform best in predicting rare recessionary events.

Table 5.4a: Forecasting Events: Returns to Fair Bet with Symmetric Payoffs, 1991q2 – 2013q2 (Post-revision outcome) (Actual Return for model $\overline{M}_{0,-1,\tau}$ ; Improvement over $\overline{M}_{0,-1,\tau}$ for other models)								
	RMSE Weights				Log Score Weights			
	$\overline{M}_{0,-1,\tau}$	$\overline{M}_{R,-1,\tau}$	$\overline{M}_{0,F,\tau}$	$\overline{M}_{R,F,\tau}$	$\overline{M}_{0,-1,\tau}$	$\overline{M}_{R,-1,\tau}$	$\overline{M}_{0,F,\tau}$	$\overline{M}_{R,F,\tau}$
<i>DROP1</i>	3.65	<b>3.43</b>	<b>3.43</b>	<b>3.43</b>	5.94	0.00	0.00	<b>1.14</b>
<i>DROP2</i>	<i>3.98</i>	<b>1.09</b>	<b>1.09</b>	<b>1.09</b>	<i>3.98</i>	<b>1.09</b>	<b>1.09</b>	<b>1.09</b>
<i>BPEAK</i>	33.78	<b>3.07</b>	1.53	0.00	<b>35.32</b>	-4.60	-3.07	-3.07
<i>BTREND</i>	8.89	-7.99	0.00	<b>4.00</b>	16.88	-7.99	<b>5.99</b>	-7.99
<i>TURN</i>	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00
<i>IRRISE</i>	18.88	11.98	<b>13.98</b>	3.99	14.88	9.99	13.98	<b>17.98</b>

Table 5.4b: Forecasting Events: Returns to Fair Bet with Asymmetric Payoffs, 1991q2 – 2013q2 (Post-revision outcome) (Actual Return for model $\overline{M}_{0,-1,\tau}$ ; Improvement over $\overline{M}_{0,-1,\tau}$ for other models)								
	RMSE Weights				Log Score Weights			
	$\overline{M}_{0,-1,\tau}$	$\overline{M}_{R,-1,\tau}$	$\overline{M}_{0,F,\tau}$	$\overline{M}_{R,F,\tau}$	$\overline{M}_{0,-1,\tau}$	$\overline{M}_{R,-1,\tau}$	$\overline{M}_{0,F,\tau}$	$\overline{M}_{R,F,\tau}$
<i>DROP1</i>	-47.17	36.33	<b>61.17</b>	<b>61.17</b>	-9.83	17.83	<b>27.83</b>	24.83
<i>DROP2</i>	18.25	47.75	41.50	<b>52.75</b>	59.75	-1.75	0.00	<b>8.25</b>
<i>BPEAK</i>	<i>28.85</i>	<b>11.25</b>	10.90	3.35	<i>26.30</i>	8.80	<b>10.45</b>	3.90
<i>BTREND</i>	4.60	-2.79	-1.86	<b>2.35</b>	7.23	-1.13	<b>0.65</b>	-3.42
<i>TURN</i>	<i>-13.50</i>	24.50	<b>28.50</b>	24.50	<i>4.00</i>	6.00	<b>15.00</b>	5.00
<i>IRRISE</i>	6.30	8.67	8.67	<b>8.74</b>	7.91	<b>7.07</b>	6.93	7.00

Note: The events are described in notes to Table 5.2a. The actual return is reported for model  $\overline{M}_{0,-1,\tau}$  and the improvement in return over  $\overline{M}_{0,-1,\tau}$  is reported for other models. Emboldened figures show the largest return.

## 5.4 Structural breaks and the usefulness of real-time data

Among others, Pesaran and Timmermann (2007) and Pesaran, Pick and Pranovich (2011) deal with forecast evaluation in the presence of regime or structural break uncertainty. Following the suggestion of Pesaran and Timmermann (2007), the empirical exercise described above can also be extended to include additional models defined using different sample periods as well as using more or less of the real-time data. This allows the combined model to trade off the advantage of extra precision on parameter estimates gained from longer samples of data against the danger of using samples that include structural breaks. Hendry and Clements (2002) suggest that the existence of structural breaks can change the ordering of models based on their forecasting performance, implying that model averaging is particularly beneficial in this case. The finding of Stock and Watson (2006) reinforces Hendry and Clements (2002) conclusions, showing that combinations of forecasts through model averaging can results in better forecasts than the ones delivered by

individual unstable models.

In this section we extend the analysis to explicitly allow for the possibility of structural breaks in the models of interest and, consequently, in the usefulness of real-time data. We adopt the meta-analysis, which is also used in chapter 4, and employ the model averaging technique in a recursive setup, allowing for the possibility that the relevant sample characterising relationships may change. In particular, all models are estimated over different sample sizes at each point in time while, through statistical forecast evaluation criteria, a weight scheme assesses whether a given estimated model still holds over time or alternatively, whether it experiences a structural break. The weights employed to obtain the "meta-model" are determined according to the forecasting performance of individual models, as well as on the possibility of a structural break. The fact that weights change over time allows us to gauge the extent to which a given model that uses revisions and/or survey data is stable over time while it provides a useful structure with which we can interpret the changing usefulness of revisions and survey data in forecasting.

Work by Pesaran and Timmermann (2007), Clark and McCracken (2010) and Giacomini and Rossi (2010) is particularly relevant here. The former suggest a combination of forecasts from models that are estimated over different sample windows using weights that are proportional to the inverse of out-of-sample loss are likely to deliver better forecasting performance if the breaks are small. Since considerable uncertainty regarding the existence of breaks is often present in applied work, this strategy provides a means of risk mitigation that is likely to result in more accurate combined forecasts. This strategy becomes especially advantageous when the breaks are small and difficult to detect since it avoids the need to locate the break point.

Moreover, the meta exercise allows us to consider the extent to which the averaging technique employed in the previous section is sufficiently flexible to incorporate the changing conditions that characterise the usefulness of revisions and survey data and whether the averaging technique is adequately robust against structural breaks. In principle, there are two possibilities:

- The averaging technique, applied in the previous section, accommodates different types of structural innovation changes and the weight scheme adopted does a good job in shifting from one model to another over time when actually, a structural break takes place;



- The averaging technique is not robust against structural breaks.

If the meta analysis conveys the same messages with the averaging technique, it means that the averaging process is sufficiently flexible to characterise data well, even in the presence of structural breaks. The estimation in rolling windows, together with the appropriate weight structure may explain why the averaging approach captures any changes in the usefulness of revisions and survey data. Since the averaging technique is much less cumbersome than the meta analysis, if the averaging technique is robust enough against breaks, the researcher is safe to use the averaging technique. But whether the averaging technique is sufficiently flexible, remains to be seen.

#### 5.4.1 Introducing structural breaks in the real-assessment

Our assessment of the usefulness of real-time data is carried out as before but we additionally consider the changing sample period for which a given model remains relevant in forecasting. The  $p$ -order vector autoregressive model that explains data (5.1) is written as before, but is now distinguished by the sample size,  $j = j_{\min}, \dots, j_{\max}$ , over which it is estimated, in addition to the reference period,  $\tau$ . We, therefore, denote the individual model by  $M_{r,f,j,\tau}$  with the subscript ' $r, f, j, \tau$ ' highlighting that the estimated model will not only differ depending on the underlying assumptions on the number of revisions, the forecast horizon in the survey and on the estimation period but also on the sample size over which the model holds.

More concretely, we compare decision-making strategies that make full use of revisions and survey data with simpler strategies that only make partial use of the available data (as before) but now decision making strategies at time  $\tau$  will be based on forecasts from model that makes use of revisions and survey data, denoted by  $M_{r,f,j,\tau}$  for  $r = 0, \dots, R$ ;  $f = -1, \dots, F$  and  $j = j_{\min}, \dots, j_{\max}$ , where  $j$  specifies the number of observations over which the model is estimated.

Essentially, we estimate a set of models and combine these using model averaging techniques, taking into account the possibility that the sample period over which a model is relevant can change through time. We call the weighted average model of these models as the "meta model" reflecting the fact that we also average across different sample sizes. The assumption is that when there is a structural break, a model is relevant over the minimum sample size. If on the other hand, there is continuity in a model's forecasting relevance, the sample over which the model holds is augmented

my additional observations as we move through time. The evolution of the average sample size over which models that use the same number of revisions and survey horizons hold provides a straightforward means for detecting potential structural breaks.

### Real-time assessment

The construction of the meta model that makes full use of the real-time data available at time  $\tau$  is obtained by the following procedure:

- We split the currently available sample into two sub-samples: an estimation period  $t = 1, \dots, \tau - \Lambda$ ; and a training period.  $t = \tau - \Lambda + 1, \dots, \tau$ .
- For each combination of  $r = 0, \dots, R$  and  $f = -1, \dots, F$  we estimate the model  $M_{r,f,j,\tau-\Lambda}$  over different sample sizes for  $j = jmin, \dots, jmax$ . The smallest models are estimated over  $t = \tau - \Lambda - jmin, \dots, \tau - \Lambda$ . Models that are taken to hold for one additional observation are estimated over  $t = \tau - \Lambda - jmin - 1, \dots, \tau - \Lambda$  while the biggest models are estimated over the whole available sample  $t = 1, \dots, \tau - \Lambda$ , so that the biggest sample size over which models are assumed to hold is given by  $jmax$ . As a result,  $(R + 1) \times (F + 2) \times (jmax - jmin + 1)$  versions of model  $M_{r,f,j,\tau-\Lambda}$  are estimated for different sample sizes over  $t = 1, \dots, \tau - \Lambda$  and for  $r = 0, \dots, R$ ,  $f = -1, \dots, F$  and  $j = jmin, \dots, jmax$ . In other words, the models are distinguished by the use of revisions and survey data as well as by the time span over which they are assumed to hold, considered here to be in operation for  $j$  periods ending in period  $\tau - \Lambda$ . When there is a break, a new period starts afresh so that the new model starts from the minimum sample size  $jmin$ . If there is no break, then when evaluating decisions at time  $\tau - \Lambda + 1$ , the model is assumed to continue from the previous period while the period over which it holds is extended by one additional observation.
- The forecast performance of the individual models is judged by comparing the nowcast obtained by the models,  $E[\tau_{-\Lambda+1+R}y_{\tau-\Lambda} \mid M_{r,f,j,\tau-\Lambda}]$ , with the post-revision value  $\tau_{-\Lambda+1+R}y_{\tau-\Lambda}$  through the calculation of mean squared error,  $s_{r,f,j,\tau-\Lambda}^p$ , or the logarithmic score,  $s_{r,f,j,\tau-\Lambda}^d$ , as discussed in section 5.2.

- The whole forecasting exercise is repeated, moving recursively, for samples over the whole of the training period. In particular, the next  $(R+1) \times (F+2) \times (j \max - j \min + 1)$  models are estimated over one additional observation producing nowcasts for  $\tau - \Lambda + 2 + R y_{\tau - \Lambda + 1}$ . Effectively, the largest models are estimated over  $t = 1, \dots, \tau - \Lambda + 1$  while the shortest models are estimated over  $t = \tau - \Lambda - j \min, \dots, \tau - \Lambda + 1$ .  $s_{r,f,j,\tau - \Lambda + 1}^p$  or  $s_{r,f,j,\tau - \Lambda + 1}^d$  are calculated accordingly. The process continues until the final  $(R+1) \times (F+2) \times (j \max - j \min + 1)$  models are estimated producing the post-revision nowcasts of  $\tau + 1 + R y_{\tau}$ . The the largest models are estimated over  $t = 1, \dots, \tau$  while the shortest models are estimated over  $t = \tau - \Lambda - j \min, \dots, \tau$ .
- The weights for the models,  $w_{r,f,j,\tau}$ , are calculated on the basis of forecast performance of the individual models over the training period. The average mean squared forecast error for each model over the training period is given by  $MSE_{r,f,j,\tau} = \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} s_{r,f,j,\tau - \Lambda + \lambda}^p$  and the average of the logarithmic score over the training period, is given by  $MLS_{r,f,j,\tau} = \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} s_{r,f,j,\tau - \Lambda + \lambda}^d$ . If there is a break, the new regime is based on a model that uses the minimum sample size, while if there is regime continuity, a model becomes relevant for one additional observation in the sample, so that the sample size over which it holds is augmented by one. The weights under the RMSE weight scheme are, then given by:

$$w_{r,f,j,\tau} = \begin{cases} \frac{\Pr(break)}{(R+1) \times (F+2)} & \text{for } j = j \min \\ w_{r,f,j-1,\tau - \Lambda - 1} \times \frac{(\sqrt{MSE_{r,f,j,\tau}})^{-1}}{\sum_r \sum_f \sum_j (\sqrt{MSE_{r,f,j,\tau}})^{-1}} \times (1 - \Pr(break)) & \text{for all } j \neq j \min \end{cases} \quad (5.15)$$

where  $\Pr(break)$  denotes the probability of break which is chosen a priori.

When the logarithmic-score is used as the statistical criterion that determines a model's nowcast accuracy, the weights are given by:

$$w_{r,f,j,\tau} = \begin{cases} \frac{\Pr(break)}{(R+1) \times (F+2)} & \text{for } j = j \min \\ w_{r,f,j-1,\tau - \Lambda - 1} \times \frac{\exp(MLS_{r,f,j,\tau})}{\sum_r \sum_f \sum_j \exp(MLS_{r,f,j,\tau})} \times (1 - \Pr(break)) & \text{for all } j \neq j \min \end{cases} \quad (5.16)$$

According to this weight structure, a model's weight depends on the model's relative nowcasting performance based on mean squared error or logarithmic score, on last period's weights,

$w_{r,f,j-1,\tau-\Lambda-1}$ , and on the probability of observing a break in the sample. The justification of including the probability of a break is that irrespective of the time series behaviour of the variables involved, there is a constant probability of a break, and if there is a break, the model enters a new regime starting again with the minimum sample size. If a previously estimated model is still relevant in forecasting, i.e. there is no break, the model just gets bigger by one additional observation as we move through the sample. At the same time, the weights are updated recursively in each period in time to reflect the extent to which different model are relevant in forecasting. If a new regime is born, such that a new model now explains the usefulness of revisions and survey data, then, the weights for each new models are given by the probability of break divided by  $(R+1) \times (F+2)$  since there are equal chances that the new regime is characterized by each of the individually estimated models,  $M_{r,f,j \min, \tau}$ , for  $r = 0, \dots, R$  and  $f = -1, \dots, F$ . The models' weights for the first period are assumed to be equal across all models.

The meta model explaining the data over the period  $t = 1, \dots, \tau$  then consists of the individually estimated models and their weights and it is denoted by:

$$\bar{M}_{R,F,J,\tau} = \{M_{r,f,j,\tau}, w_{r,f,j,\tau} \text{ for } r = 0, \dots, R, f = -1, \dots, F \text{ and } j = jmin, \dots, jmax\}. \quad (5.17)$$

The meta model can be used to obtain point forecasts and density forecasts using the weighted average of the models' individual point forecasts and aggregating over the models' individual densities. The statistics that follow provide a clear characterisation of the nature of the meta models obtained at time  $\tau$ :

$$\tilde{\mu}_{\tau}^r = \sum_{r=0}^R r \times w_{r,f,j,\tau} \quad (5.18)$$

$$\tilde{\mu}_{\tau}^f = \sum_{f=-1}^F f \times w_{r,f,j,\tau} \quad (5.19)$$

$$\tilde{\mu}_{\tau}^j = \sum_{j=j \min}^{j \max} j \times w_{r,f,j,\tau}. \quad (5.20)$$

As with the average models, the  $\tilde{\mu}_{\tau}^r$  and  $\tilde{\mu}_{\tau}^f$  statistics capture the relative importance of the revision data and survey data in defining the meta model at time  $\tau$ . Any deviation from 0 and  $-1$  respectively shows that revision and survey data make a contribution to the out-of-sample forecasting exercise. The statistic  $\tilde{\mu}_{\tau}^j$  shows the weighted average sample size over which the meta model holds. A rising

$\tilde{\mu}_\tau^j$  over time is a sign of regime continuity and that a particular meta model continues to be relevant in forecasting. On the other hand, if a structural break has taken place then it should be reflected as a decline in the  $\tilde{\mu}_\tau^j$  statistic. In other words, if the weighted average sample length drops to low levels, then the usefulness of revision and survey data has changed compared to previous periods. The evolution of  $\tilde{\mu}_\tau^j$  statistic across time and different economic phases reflects the changes in the usefulness of revision and survey data sometimes occurring abruptly and sometimes gradually. Taken together, the three statistics provide a straightforward real-time assessment of the usefulness of the revision and survey data in the meta model  $\bar{M}_{R,F,J,\tau}$  at different points in time.

### Final assessment

As before, the final assessment entails judging the usefulness of real time data based on models' ability to predict a set of economic events involving output and inflation. Using a decision-based approach we compare the nowcasting performance of decision-making strategies that make full use of the revision and survey data to simpler strategies that make only partial use of the data. Accordingly, the final assessment compares strategies based on the all-inclusive meta model  $\bar{M}_{R,F,J,\tau}$  to simpler strategies that make use of models without survey data ( $\bar{M}_{R,-1,J,\tau}$ ), or without revisions data ( $\bar{M}_{0,F,J,\tau}$ ) and even the model that does not incorporate either revisions or survey data (i.e. the conventional real-time model  $\bar{M}_{0,-1,J,\tau}$ ).

In terms of statistical evaluation criteria, the nowcasts/forecasts obtained from meta models are judged according to MSE or the logarithmic score, as described in section 5.2. Most importantly, the judgement on the usefulness of real time data involves economic evaluation criteria that reflect the associated returns from predicting rare events in an explicit investment scenario. The events considered are the same as in section 5.2, namely, DROP1, DROP2, BPBREAK, BTREND, TURN and IRRISE. The probability of these events are nowcast using the meta models  $\bar{M}_{R,F,J,\tau}$ ,  $\bar{M}_{0,F,J,\tau}$ ,  $\bar{M}_{R,-1,J,\tau}$  and  $\bar{M}_{0,-1,J,\tau}$  delivering hit rates and Kuipers scores as given by (5.10) and (5.11) respectively. The returns associated with explicit investment scenarios are computed for each meta model,  $\bar{M}_{R,F,J,\tau}$ ,  $\bar{M}_{0,F,J,\tau}$ ,  $\bar{M}_{R,-1,J,\tau}$  and  $\bar{M}_{0,-1,J,\tau}$ , according to the procedure outlined in section 5.2 and based on equations (5.12)- (5.14).

### 5.4.2 Results

Employing the same real time dataset as in section 5.3, we make additional modelling choices. In particular, we set the probability of break,  $\Pr(break)$ , equal to 0.01. Robustness check has shown that the choice of the probability of break does not affect the results. Nonetheless, the choice of 0.01 for the probability of break was made in order to reflect the results from structural break tests applied to the series involved. While most of the series experienced no break and their time series properties remained stable over time, some series exhibited some form of instability exhibiting two or three breaks over the 184-quarter period 1968q4 – 1914q3. Taken probabilistically, this means that there is almost 0.01 chance of observing a break in the sample.

The second modelling choice we make is that we set  $j_{\min} = 45$  quarters, noting that the choice of the minimum regime length is driven by the need to have enough observations for estimation purposes. To be consistent with the exercise performed based on the model averaging technique (section 5.3), we set the maximum period for the survival of unchanged usefulness of survey and revisions data equal to  $j_{\max} = 80$  quarters while we maintain the assumption that output revisions continue no longer than three periods after the first release and investors make use of survey data up to three periods ahead i.e.  $r = 0, 1, 2, 3$  and  $f = -1, 0, 1, 2, 3$ . Thus, for every point in time, there are  $80 - 45 + 1 = 36$  candidate sample sizes over which individual models are likely to hold. Given this setup, there are  $(j_{\max} - j_{\min} + 1) * (R + 1) * (F + 2) = 36 * 4 * 5 = 720$  versions of the model  $M_{r,f,j,\tau}$  for  $r = 0, 1, 2, 3$ ,  $f = -1, 0, 1, 2, 3$  and  $j = 45, \dots, 80$ .

#### Real-time forecast evaluation of point and density forecasts

The models' weights for the first period are assumed to be equal across all models, so that  $w_{r,f,j,1991q2} = 1/720$  for all  $r = 0, 1, 2, 3$ ,  $f = -1, 0, 1, 2, 3$  and  $j = 45, \dots, 80$ . Thus, the average sample size in the first set of results is equal to  $\frac{45+80}{2} = 62.5$ . The next step involves estimating the meta model  $\overline{M}_{R,F,1991q3}$  using the real time data available for 1969q1 – 1991q3 and using the 80-quarter period 1969q1 – 1988q4 in estimation while the training period holds from 1988q4 to 1991q3. Twenty versions of the model  $M_{r,f,j_{\max},1988q4}$  for  $r = 0, 1, 2, 3$  and  $f = -1, 0, 1, 2, 3$  are estimated over the 80-quarter period 1969q1 – 1988q4. Twenty versions of model  $M_{r,f,j_{\max}-1,1988q4}$  for  $r = 0, 1, 2, 3$  and  $f = -1, 0, 1, 2, 3$  are estimated over the 79-quarter period 1969q2 – 1988q4 and

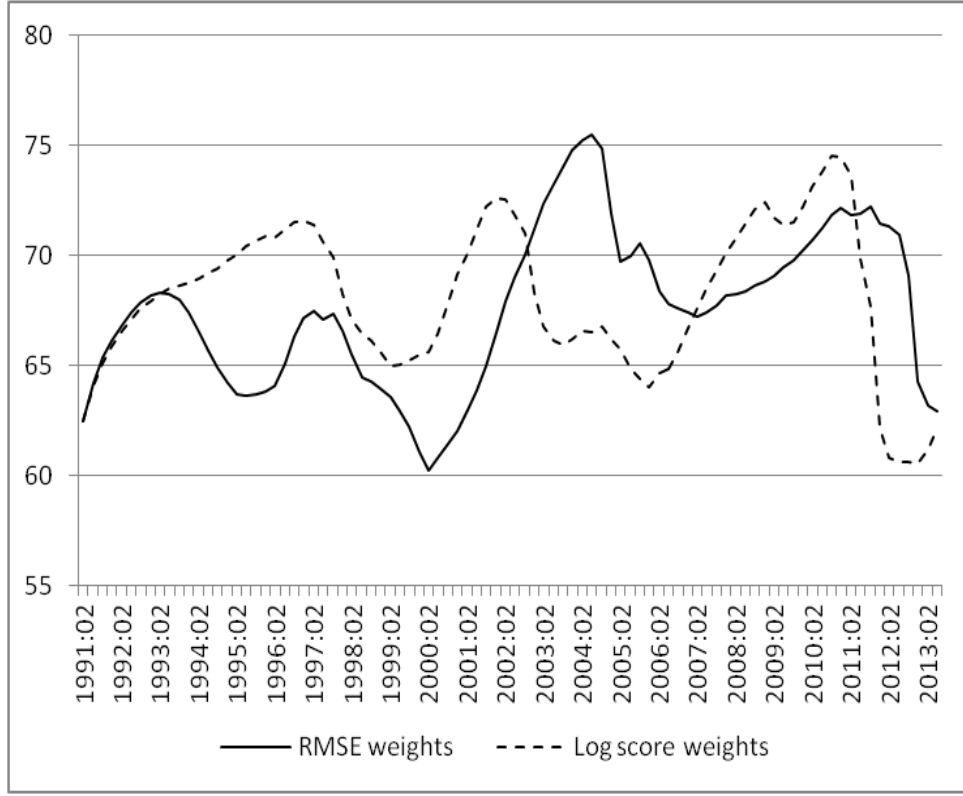


Figure 5-8: The weighted average sample size,  $\tilde{\mu}_\tau^j$ , based on meta-modelling.

so on until the smallest models  $M_{r,f,j \min, 1988q4}$  for  $r = 0, 1, 2, 3$  and  $f = -1, 0, 1, 2, 3$  are estimated over the 45-quarter period  $1977q4 - 1988q4$ . Each of these 720 models are used to produce the four-period ahead forecast of post-revision measure of contemporaneous output and this is compared to the observed outcome,  $1989q4y_{1988q4}$ . We then move through the training period: The largest models are now estimated over the 81-quarter period  $1969q1 - 1989q1$  while the smallest models are estimated over the 46-quarter period  $1977q4 - 1989q1$ . Each of the 720 are used to predict the post-revision nowcast,  $1990q1y_{1989q1}$ . This is repeated over the whole training period moving recursively and judging the relative performance of the 720 models each time to obtain the set of weights,  $w_{r,f,1991q3}$ , defined in (5.15) for the RMSE and (5.16) for the log scores. Moving on one period, this entire exercise can be repeated over the sample  $1969q2 - 1991q4$ , using  $1989q1 - 1991q4$  as the training period, to derive  $w_{r,f,1991q4}$ .

Figure 5.8 shows the weighted average sample size,  $\tilde{\mu}_\tau^j$ , given by (5.20), under the logarithmic score and RMSE weight schemes. The figure demonstrates that, with very few exceptions,

the weighted average sample size fluctuated above 62.5 observations which is the average sample considered in the meta approach. (Remember  $j_{\min} = 45$  and  $j_{\max} = 80$ ). This clearly provides evidence that most of the time, the weight structure allocated most of the weights to the larger models. In many cases, the weighted average sample size exceeded 70 observations and was very close to the maximum of 80 observations, which is the constant sample size considered in the averaging technique presented in the previous section.

Our findings echo results by Pesaran and Timmermann (2007) who demonstrate that pre-break data, which one might be tempted to characterise as irrelevant or misleading for the forecasting exercise, may actually contain useful information to such an extent that it can be optimal to include them in the estimation window. This tactic is shown to deliver forecasts that have lower variance. The fact that the meta technique outcomes call for larger models indicates that benefits from increased precision of parameter estimates due to larger estimation window overshadow the benefits from using small estimation samples to account for the possibility of breaks. The model averaging technique gives important insights about the trade-off between bias and forecast error variance as outlined in Pesaran and Timmermann (2007) since the exercise has shown that the forecasting performance is enhanced when models are estimated over longer sample sizes rather than using post-break data only.

Figure 5.9 shows  $\tilde{\mu}_{\tau}^r$ , defined in (5.18) i.e. the weighted average of the revision length of the models included in the meta model  $\overline{M}_{R,F,J,\tau}$ . The weighted average revision length for the first period in 1991q2, is exactly equal to 1.5 which is mid-way between zero and three, the minimum and maximum values since it was specified that all models received equal weights in the first period. Unlike the case of average models, the plots under the two weight schemes move together and there is not too much difference between the message conveyed by each of them. The plots based on RMSE and log score illustrate that while the average revisions horizon remained stable at just below 1 until 2001q3, the weighted average revision length increased substantially from then on. In the years after the financial crises the weighted average revision length reached the maximum of three using the log score as the criterion, while the statistic based on RMSE weights also reached 2.5. This matches the findings from figure 5.2 which demonstrates that revisions become more important towards the end of the sample: In the early 2000s and in the years after the financial



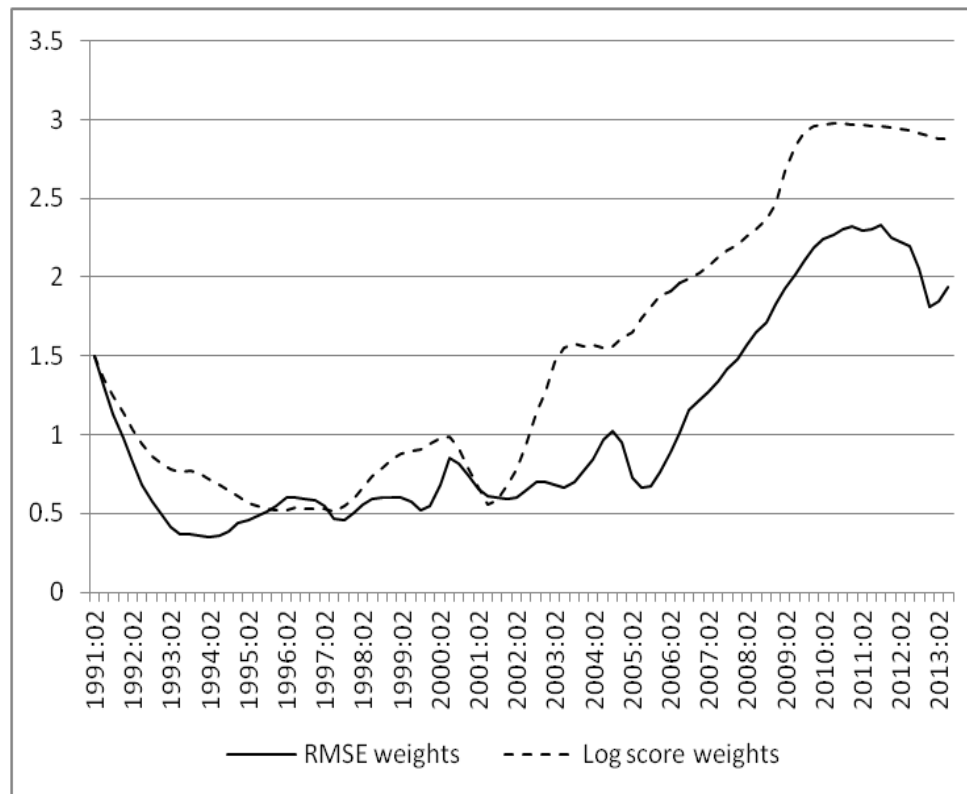


Figure 5-9: The weighted average use of revisions data in post revision output forecasts,  $\tilde{\mu}_T^r$ , based on meta-modelling.

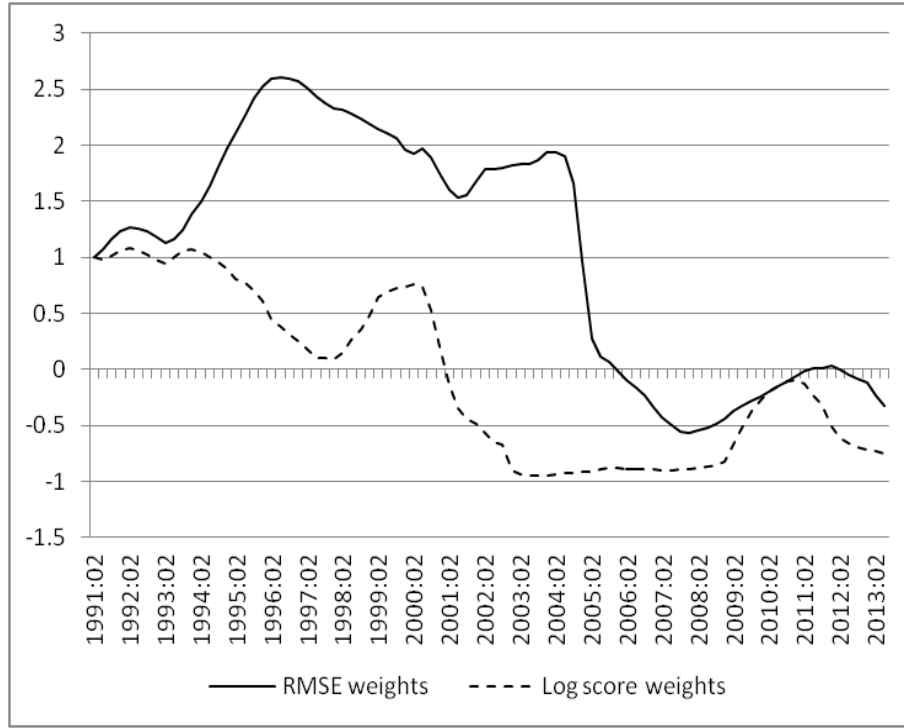


Figure 5-10: The weighted average use of revisions data in post revision output forecasts,  $\tilde{\mu}_\tau^f$ , based on meta-modelling.

crisis there were a number of large revisions in the output data so that the meta model adjusts to exploit the extra information in revisions.

Figure 5.10 provides the evolution of the weights average survey horizon,  $\tilde{\mu}_\tau^f$ , defined in (5.19). Again, the statistic starts from 1, mid-way between the maximum and minimum horizon considered, as we set the first period's weights to be equal across models. Although the statistics based RMSE and logarithmic score weights follow different directions over the start of the sample, the bigger picture is that the usefulness of expectations data declines over time, indicating that the inclusion of survey data did not enhance the ability of individual models to predict the post-revision outcome of output. As the graph suggests, after 2009q3, the average survey horizon hardly exceeds zero implying, that models that make no use of survey data received higher weights.

Although the average survey horizon under the averaging technique (figure 5.3) remains above zero throughout the sample, the two approaches (averaging and meta) collectively emphasise that in the real-assessment, the role of survey data in forecasting is undermined as we move through

time.

### Final-assessment of model predictability based on economic events

The final assessment is carried out as before, albeit based on the meta models obtained by accommodating the possibility of structural breaks in the relationships of interest. The meta models, that differ in their use of real-time data and the sample size over which they hold, compete themselves in predicting the occurrence of six events, as outlined above.

Table 5.5 shows the results based on the GW test of equal forecasting performance. The conventional meta model  $\overline{M}_{0,-1,J,\tau}$  has an average RMSE of 1.11% over the whole evaluation period with the three meta models  $\overline{M}_{R,F,J,\tau}$ ,  $\overline{M}_{R,-1,J,\tau}$  and  $\overline{M}_{0,F,J,\tau}$  achieving lower RMSEs but only meta models  $\overline{M}_{R,F,J,\tau}$  and  $\overline{M}_{R,-1,J,\tau}$  show a statistically significant improvement. The conventional model  $\overline{M}_{0,-1,J,\tau}$  delivers an average log score of  $-2.56$  while the meta models  $\overline{M}_{R,F,J,\tau}$  and  $\overline{M}_{R,-1,J,\tau}$  outperform it achieving higher values. However, the improvement is not statistically significant. The meta model  $\overline{M}_{0,F,J,\tau}$  is shown to deteriorate in forecast performance compared to  $\overline{M}_{0,-1,J,\tau}$  but the deterioration is not statistically significant. The test results suggest that incorporating real-time data can improve the forecasting of output although the argument for the use of revisions is more compelling than for the use of surveys, a result that also prevails in section 5.2 where the analysis was based on average models.

Table 5.5: RMSE and Average Logarithmic Predictive Scores for Output Growth Nowcasts:

1991q2 – 2013q2 (Post-revision Outcome)

*Actual RMSE and Average Logarithmic scores for model  $\overline{M}_{0,-1,J,\tau}$*

*Scaled difference from model  $\overline{M}_{0,-1,J,\tau}$  for other models*

Average Model	RMSE	Log-score
	Actual for $\overline{M}_{0,-1,J,\tau}$	
$\overline{M}_{0,-1,J,\tau}$	0.0111	-2.326
	Relative to $\overline{M}_{0,-1,J,\tau}$	
$\overline{M}_{R,-1,J,\tau}$	-0.0003***	0.4731
$\overline{M}_{0,F,J,\tau}$	-0.0004	-0.1533
$\overline{M}_{R,F,J,\tau}$	-0.0002**	0.4586

Note: The meta model  $\overline{M}_{R,F,J,\tau}$  is as defined in (5.17). In particular, model  $\overline{M}_{R,F,J,\tau}$  denotes the unrestricted meta model which includes use of the full available revisions and survey data on expectations for  $\tau = \underline{\tau}, \dots, T$  estimated over different sample sizes;  $\overline{M}_{0,F,J,\tau}$  denotes the model where no use of the revisions data is made for  $\tau = \underline{\tau}, \dots, T$  and is estimated over different sample sizes;  $\overline{M}_{R,-1,J,\tau}$  denotes the model where no use is made of the survey data for  $\tau = \underline{\tau}, \dots, T$  and is estimated over different sample sizes while  $\overline{M}_{0,-1,J,\tau}$  denotes the model where no use is made of either the revisions or survey data throughout (i.e. based on the conventional real time model), for  $\tau = \underline{\tau}, \dots, T$ , estimated over different sample sizes. A ‘\*’ denotes significance at the 10% level, ‘\*\*’ denotes significance at 5% level and ‘\*\*\*’ significance at the 1% level of the Giacomini-White (2006) test of equal forecast performance testing whether the RMSE and the logarithmic predictive score are significantly smaller or larger, respectively, than the corresponding statistics from model  $\overline{M}_{0,-1,J,\tau}$ .

**Evaluation of event probabilities and fair bet outcomes** Tables 5.6a and 5.6b summarise the degree of correspondence between model’s prediction of the occurrence of events and the actual realisation of events. In particular, table 5.6a shows the proportion of forecasts that are correct, as shown in (5.10) for each meta model and Table 5.6b reports the Kuipers scores, as calculated by (5.11). Table 5.6b also reports, in parentheses, the results of the static and dynamic versions of Pesaran and Timmermann (2009) tests of the null that a model’s forecast performance is no better than what would have been achieved if predictions were made randomly. When the RMSE is the criterion, it appears that the meta model  $\overline{M}_{0,F,J,\tau}$  achieves the highest hit rates in most cases. The other meta models,  $\overline{M}_{R,F,J,\tau}$ ,  $\overline{M}_{R,-1,J,\tau}$  and  $\overline{M}_{0,-1,J,\tau}$ , achieve the highest hit rates in certain occasions but their superiority is only marginal. For instance,  $\overline{M}_{0,-1,J,\tau}$  achieves a high hit rate when predicting event DROP1 but it appears that other models perform equally well attaining the same value. Another example is the prediction of event IRRISE. The conventional model  $\overline{M}_{0,-1,J,\tau}$  achieves the highest hit rate together with model  $\overline{M}_{R,F,J,\tau}$  but the improvement over the other meta models is only minimal. When log score weights are used, the meta models  $\overline{M}_{R,-1,J,\tau}$  and  $\overline{M}_{0,F,J,\tau}$  appear to achieve the highest hit rates in most cases.

The pattern of results based on the Kuipers score is less easy to interpret. The meta models are not particularly successful in predicting relatively frequent events such as BTREND which is

as likely to happen as not. Accordingly the Pesaran and Timmermann tests indicate that models' predictions are not better than what would have been achieved if predictions were random. It appears that models' performance is very similar when predicting BTREND and IRRISE. In addition, none of the meta models is good in predicting the very rare event TURN. All models achieve zero Kuipers score. However, under the RMSE weight scheme, the meta model  $\overline{M}_{R,-1,J,\tau}$  achieves the highest Kuipers score when predicting DROP1 and DROP2 and appears to be somewhat more dominant. Similar conclusions are drawn based on log score weights.

Table 5.6a: Forecasting Events: Hit Rate: 1991q2 – 2013q2 (Post-revision Outcome)									
	$p_u$	RMSE Weights				Log Score Weights			
		$\overline{M}_{0,-1,J,\tau}$	$\overline{M}_{R,-1,J,\tau}$	$\overline{M}_{0,F,J,\tau}$	$\overline{M}_{R,F,J,\tau}$	$\overline{M}_{0,-1,J,\tau}$	$\overline{M}_{R,-1,J,\tau}$	$\overline{M}_{0,F,J,\tau}$	$\overline{M}_{R,F,J,\tau}$
<i>DROP1</i>	7%	<b>0.933</b>	<b>0.933</b>	<b>0.933</b>	0.921	<b>0.933</b>	0.921	<b>0.933</b>	<b>0.933</b>
<i>DROP2</i>	5%	0.955	<b>0.966</b>	0.955	0.955	0.955	<b>0.966</b>	0.955	0.955
<i>BPEAK</i>	22%	0.899	0.888	<b>0.910</b>	<b>0.910</b>	0.910	0.876	<b>0.921</b>	0.910
<i>BTREND</i>	48%	0.528	0.517	<b>0.573</b>	0.517	<b>0.584</b>	<b>0.584</b>	<b>0.584</b>	0.506
<i>TURN</i>	2%	0.978	0.978	0.978	0.978	0.978	0.978	0.978	0.978
<i>IRRISE</i>	48%	<b>0.618</b>	0.607	0.607	<b>0.618</b>	0.607	<b>0.652</b>	0.607	0.629

Note: The meta model  $\overline{M}_{R,F,J,\tau}$  is as defined in (5.17). Event DROP1 is one-period negative output growth; DROP2 is two successive periods of negative output growth; BPEAK is output level below previous peak; BTREND is output level below 5-period moving average; TURN is a turning point in output; IRRISE is interest rate rise by Taylor rule.  $p_u$  is the unconditional probability of the event for the period 1991q2 – 2013q2. Emboldened figures show the largest hit rate.

Table 5.6b: Forecasting Events: Kuipers Score: 1991q2 – 2013q2 (Post-revision Outcome)									
	$p_u$	RMSE Weights				Log Score Weights			
		$\overline{M}_{0,-1,J,\tau}$	$\overline{M}_{R,-1,J,\tau}$	$\overline{M}_{0,F,J,\tau}$	$\overline{M}_{R,F,J,\tau}$	$\overline{M}_{0,-1,J,\tau}$	$\overline{M}_{R,-1,J,\tau}$	$\overline{M}_{0,F,J,\tau}$	$\overline{M}_{R,F,J,\tau}$
<i>DROP1</i>	7%	0.000 (–,–)	<b>0.155</b> (**,–)	0.000 (–,–)	–0.012 (–,**)	0.000 (–,–)	<b>0.142</b> (*,–)	0.000 (–,–)	<b>0.000</b> (–,–)
<i>DROP2</i>	5%	0.000 (–,–)	<b>0.250</b> (***,**)	0.000 (–,–)	0.000 (–,–)	0.000 (–,–)	<b>0.250</b> (***,**)	0.000 (–,–)	0.000 (–,–)
<i>BPEAK</i>	22%	0.550 (***,–)	0.536 (***,**)	0.636 (***,**)	<b>0.671</b> (***,*)	0.600 (***,–)	0.557 (***,–)	0.650 (***,**)	<b>0.671</b> (***,*)
<i>BTREND</i>	48%	0.060 (–,–)	0.039 (–,–)	<b>0.156</b> (–,–)	0.039 (–,–)	0.174 (–,–)	0.173 (–,*)	<b>0.176</b> (–,–)	0.013 (–,–)
<i>TURN</i>	2%	0.000 (–,–)	0.000 (–,–)	0.000 (–,–)	0.000 (–,–)	0.000 (–,–)	0.000 (–,–)	0.000 (–,–)	0.000 (–,–)
<i>IRRISE</i>	48%	<b>0.244</b> (**,* )	0.218 (**,* )	0.221 (**,* )	0.243 (**,* )	0.226 (**,* )	<b>0.309</b> (***,**)	0.219 (**,* )	0.264 (***,**)

Note: The meta model  $\overline{M}_{R,F,J,\tau}$  is as defined in (5.17).  $p_u$  is the unconditional probability of the event for the period 1991q2 – 2013q2; emboldened figures show the largest Kuipers score; the figures in parentheses (a,b) below the Kuipers Scores show the outcome of the static and dynamic versions of the Pesaran and Timmermann (2009) tests of no additional predictive power beyond that of the unconditional probability; a ‘\* \* \*’ indicates significance at 1% level, ‘\*\*’ indicates significance at 5% level, ‘\*’ indicates significance at 10% level, and ‘–’ indicates no significance at 10% level.

Tables 5.7a and 5.7b summarise the results on the fair bet investment scenarios, reporting the returns achieved by the conventional model  $\overline{M}_{0,-1,J,\tau}$  and the improvement over the returns of  $\overline{M}_{0,-1,J,\tau}$  for the remaining meta models. Table 5.7a which reports results for the symmetric fair bet, shows that, under RMSE weight scheme, meta model  $\overline{M}_{0,F,J,\tau}$  achieves the highest returns in many cases while the meta model  $\overline{M}_{R,-1,J,\tau}$  achieves lower returns than the conventional model in certain occasions. When weights are based on log scores, the meta models that incorporate real-time data to a certain extent achieve higher returns than the conventional model in most cases, while each of  $\overline{M}_{0,F,J,\tau}$ ,  $\overline{M}_{R,-1,J,\tau}$  and  $\overline{M}_{R,F,J,\tau}$  attain strictly the highest returns in predicting BREAK, DROP2 and IRRISE respectively. It should be emphasised that the results are not as

dramatic as described in section 5.2 where analysis is based on average models. The inability of meta models to predict frequent or very rare events partly accounts for this discrepancy.

The results on the symmetric scenario have not been particularly enlightening, despite the fact that to a certain extent it has been shown that the predictive ability of models is enhanced by the use of real-time data. The asymmetric setup delivers higher payout on rare events and so emphasises the ability of meta models to predict more infrequent events. Using RMSE weights, the meta model  $\overline{M}_{0,F,J,\tau}$  is the dominant model, achieving the highest returns even for the most rare events such as DROP1 and TURN which have unconditional probabilities equal to 7% and 2% respectively. Moreover, model  $\overline{M}_{R,-1,J,\tau}$  appears to perform decently achieving the highest returns in two occasions. When log score weights are used, meta model  $\overline{M}_{R,F,J,\tau}$  together with meta model  $\overline{M}_{0,F,J,\tau}$  collectively achieve the highest returns for the vast majority of events. The over-riding conclusion is that models which include survey data perform best in predicting rare events, reflecting the findings from the analysis based on average models.

Taking everything into consideration, the results based on the meta approach have shown that the weight structure assigned the biggest portion of weights to meta models that were biggest in sample size, suggesting that the benefits of more accurate estimation exceed the possible advantages of accounting for structural instability. To this end, the average technique considered in section 5.2, estimated models over the largest sample size, is sufficiently flexible and reliable as it is carried out in a recursive manner so there is no need to run into cumbersome exercises in order to explicitly allow for structural breaks. In addition, the meta exercise has shown that under the RMSE weight scheme the model that delivered overall the best hit rates, Kuipers scores and profits in the asymmetric and symmetric case investment scenarios is meta model  $\overline{M}_{0,F,J,\tau}$ , while model  $\overline{M}_{0,F,J,\tau}$  together with  $\overline{M}_{R,F,J,\tau}$  are the winners under the logarithmic score weight scheme.

Table 5.7a: Forecasting Events: Returns to Fair Bet with Symmetric Payoffs, 1991q2 – 2013q2 (Post-revision Outcome) (Actual Return for model $\overline{M}_{0,-1,J,\tau}$ ; Improvement over $\overline{M}_{0,-1,J,\tau}$ for other models)								
	RMSE Weights				Log Score Weights			
	$\overline{M}_{0,-1,J,\tau}$	$\overline{M}_{R,-1,J,\tau}$	$\overline{M}_{0,F,J,\tau}$	$\overline{M}_{R,F,J,\tau}$	$\overline{M}_{0,-1,J,\tau}$	$\overline{M}_{R,-1,J,\tau}$	$\overline{M}_{0,F,J,\tau}$	$\overline{M}_{R,F,J,\tau}$
<i>DROP1</i>	<b>5.94</b>	<b>0.00</b>	<b>0.00</b>	-1.14	<b>5.94</b>	-1.14	<b>0.00</b>	<b>0.00</b>
<i>DROP2</i>	3.98	<b>1.09</b>	0.00	0.00	3.98	<b>1.09</b>	0.00	0.00
<i>BPEAK</i>	33.78	-1.54	<b>1.53</b>	<b>1.53</b>	35.32	-4.60	<b>1.53</b>	0.00
<i>BTREND</i>	4.89	-2.00	<b>8.00</b>	-2.00	<b>14.88</b>	<b>0.00</b>	<b>0.00</b>	-13.98
<i>TURN</i>	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00
<i>IRRISE</i>	<b>20.88</b>	-2.00	-2.00	<b>0.00</b>	18.88	4.00	0.00	<b>7.99</b>

Table 5.7b: Forecasting Events: Returns to Fair Bet with Asymmetric Payoffs, 1991q2 – 2013q2 (Post-revision Outcome) (Actual Return for model $\overline{M}_{0,-1,J,\tau}$ ; Improvement over $\overline{M}_{0,-1,J,\tau}$ for other models)								
	RMSE Weights				Log Score Weights			
	$\overline{M}_{0,-1,J,\tau}$	$\overline{M}_{R,-1,J,\tau}$	$\overline{M}_{0,F,J,\tau}$	$\overline{M}_{R,F,J,\tau}$	$\overline{M}_{0,-1,J,\tau}$	$\overline{M}_{R,-1,J,\tau}$	$\overline{M}_{0,F,J,\tau}$	$\overline{M}_{R,F,J,\tau}$
<i>DROP1</i>	-17.67	22.67	<b>38.67</b>	33.67	4.00	0.00	5.00	<b>8.00</b>
<i>DROP2</i>	37.50	25.50	22.25	<b>31.50</b>	58.75	-15.00	0.00	<b>11.25</b>
<i>BPEAK</i>	38.75	<b>4.35</b>	3.00	2.45	31.30	4.35	<b>7.45</b>	3.90
<i>BTREND</i>	-2.12	5.14	<b>5.28</b>	2.07	<b>9.30</b>	-2.14	-2.00	-11.49
<i>TURN</i>	<b>16.00</b>	-13.00	<b>0.00</b>	-5.00	1.00	4.00	<b>8.00</b>	<b>8.00</b>
<i>IRRISE</i>	9.37	<b>2.79</b>	0.79	0.86	9.44	2.79	2.86	<b>3.86</b>

Note: The events are described in notes to Table 5.5a. The actual return is reported for model  $\overline{M}_{0,-1,J,\tau}$  and the improvement in return over  $\overline{M}_{0,-1,J,\tau}$  is reported for other models. Emboldened figures show the largest return.



## 5.5 Conclusion

The empirical exercise provides clear-cut evidence that forecasts of output growth and recessionary events are enhanced through the use of real-time data. The ‘real-time’ and ‘final’ evaluations of the forecasts from the VAR models considered in this chapter show that point forecasts and density forecasts are improved by using survey data on expected future output movements and by using first-release and revisions data. The exercise shows that this is especially true if, as here, the modelling takes into account that the data can be more or less helpful at different times, with the revisions data appearing to be particularly important during downturns when larger (but predictable) measurement errors appear in the first-release data. On the other hand, it is the survey data which seems particularly important when forecasting the likelihood of recessionary events. These are relatively rare and extreme events which conventional linear forecasting models might struggle to accommodate but which are incorporated into professional forecasters’ predictions reasonably quickly. Survey data therefore provides the means to quickly include this information in a time series model so that, again, forecast performance is improved by allowing the data to be used more or less intensively at different times.

The chapter also uses meta analysis, which accommodates uncertainty regarding the existence of breaks, and allows for changing sample sizes as well as the changing usefulness of real time data, captured by the averaging technique. Results demonstrate that relatively few weights are given to models with shorter sample sizes implying that gains from the improved precision of parameter estimates found in long samples outweigh any benefits of explicitly capturing structural breaks. This suggests that the averaging technique used in the first part of this chapter is sufficiently flexible to accommodate the changing usefulness of data over time.

The current work in this chapter can be extended to allow model weights to vary with evaluation criteria. For example a model’s weight could depend on the fair-bet profits accumulated by an investor.

## 5.6 Appendix A

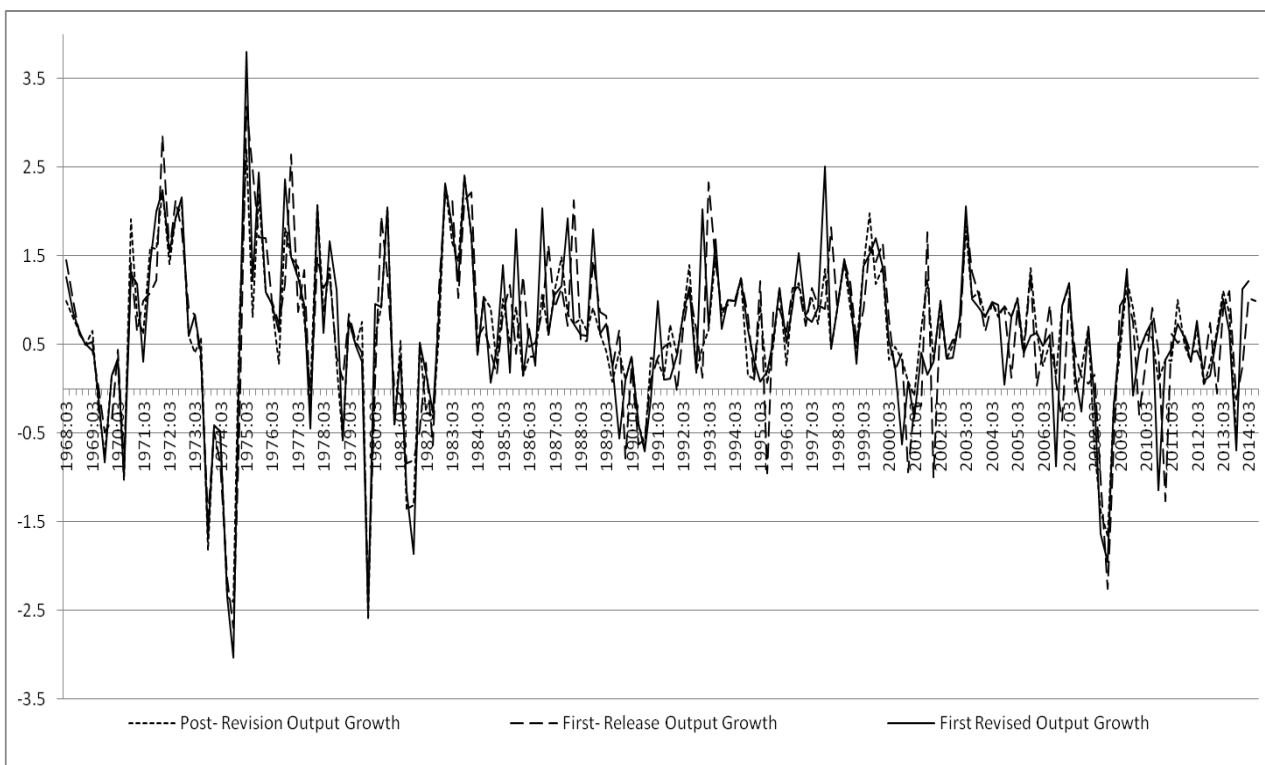


Figure A1: Post-Revision, First-Release and First-Revised Output Growth defined as  $t+4y_t - t+4y_{t-1}$  and  $t+1y_t - t+1y_{t-1}$  and  $t+2y_t - t+2y_{t-1}$ , respectively.

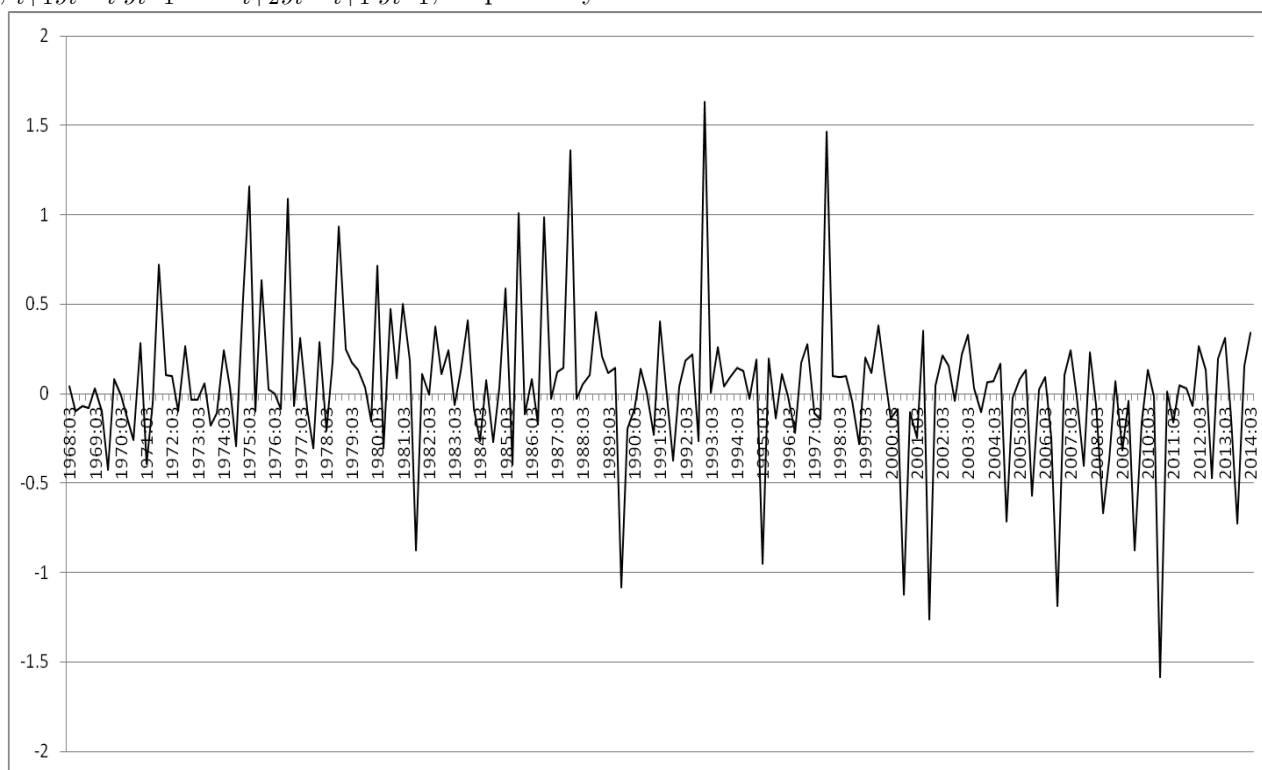


Figure A2: The first revision of the real GDP series (Percentage Change)

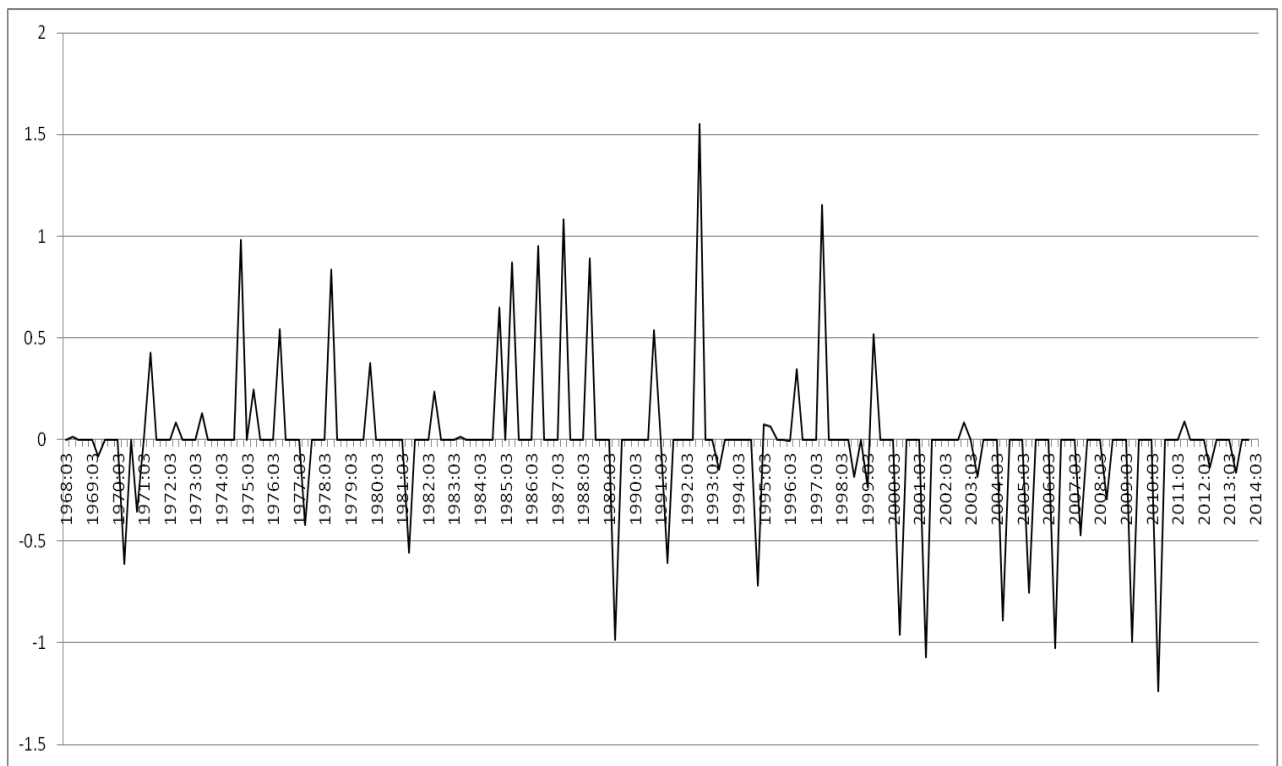


Figure A3: The second revision of the real GDP series (Percentage Change)

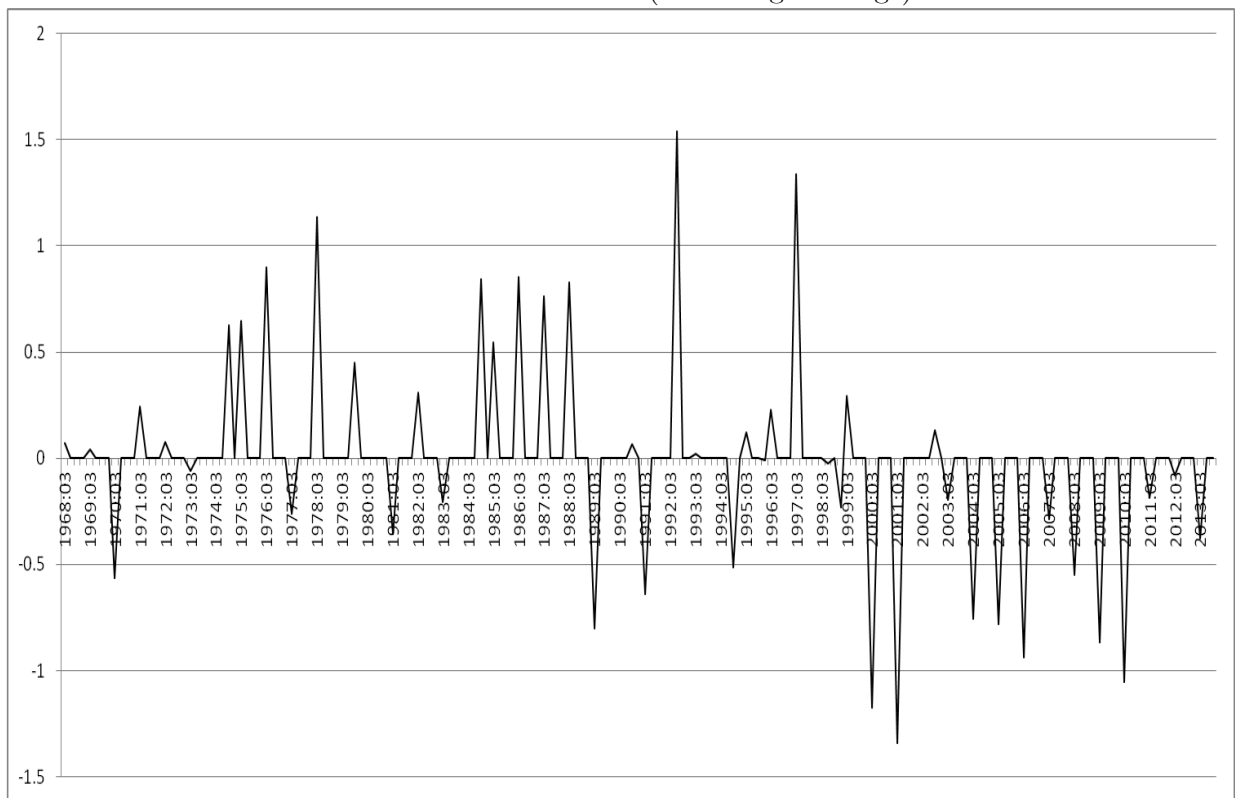


Figure A4: The third revision of the real GDP series (Percentage Change)

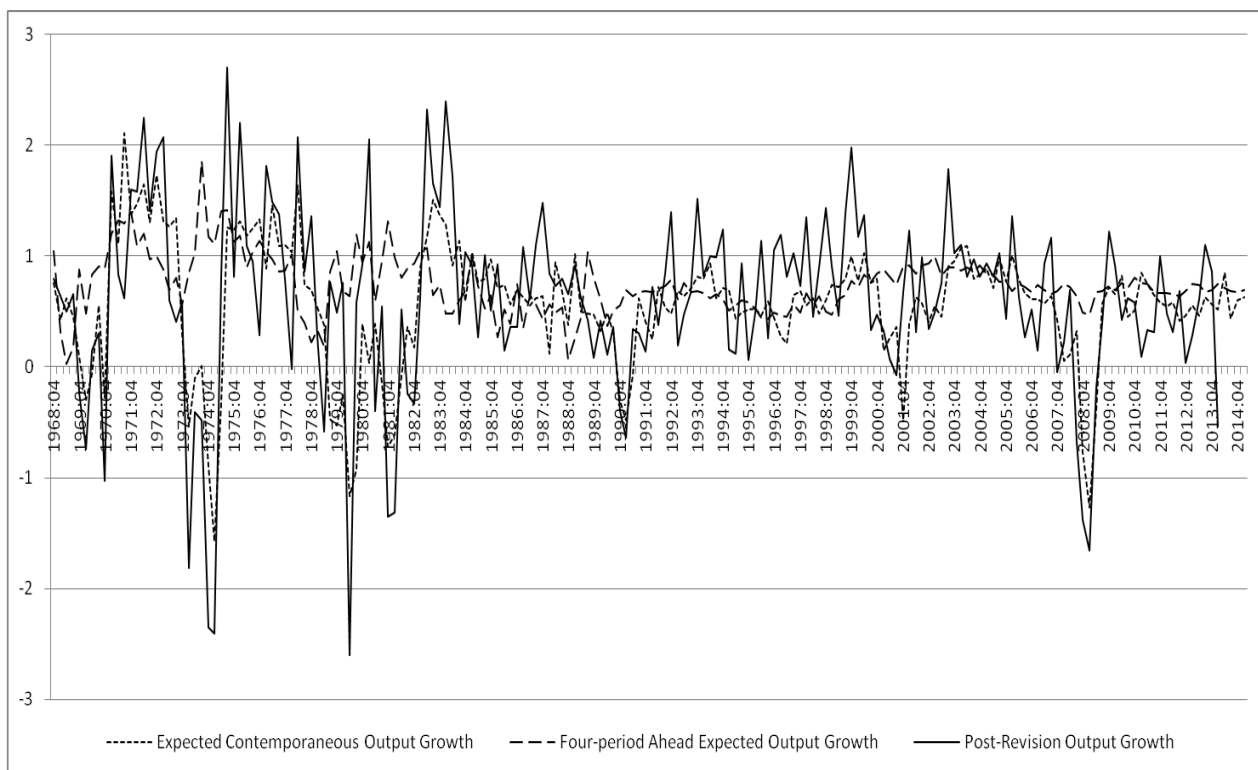


Figure A5: Expected Contemporaneous, Expected Four-Period Ahead and Post-Revision Output Growth, defined as  $y_t - y_{t-1}$ ,  $y_{t-4} - y_{t-5}$  and  $y_{t+4} - y_{t+3}$  respectively.

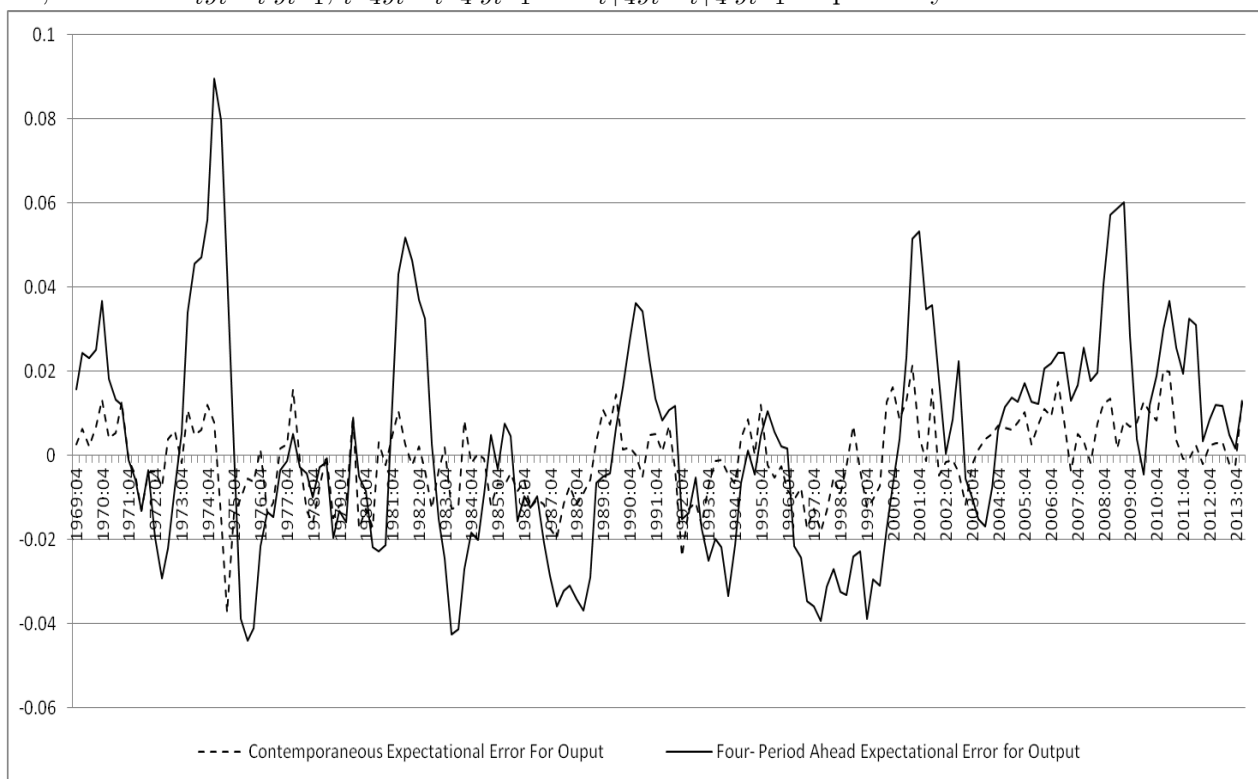


Figure A6: Contemporaneous and Four-Period Ahead Expectational Errors for the Output Series

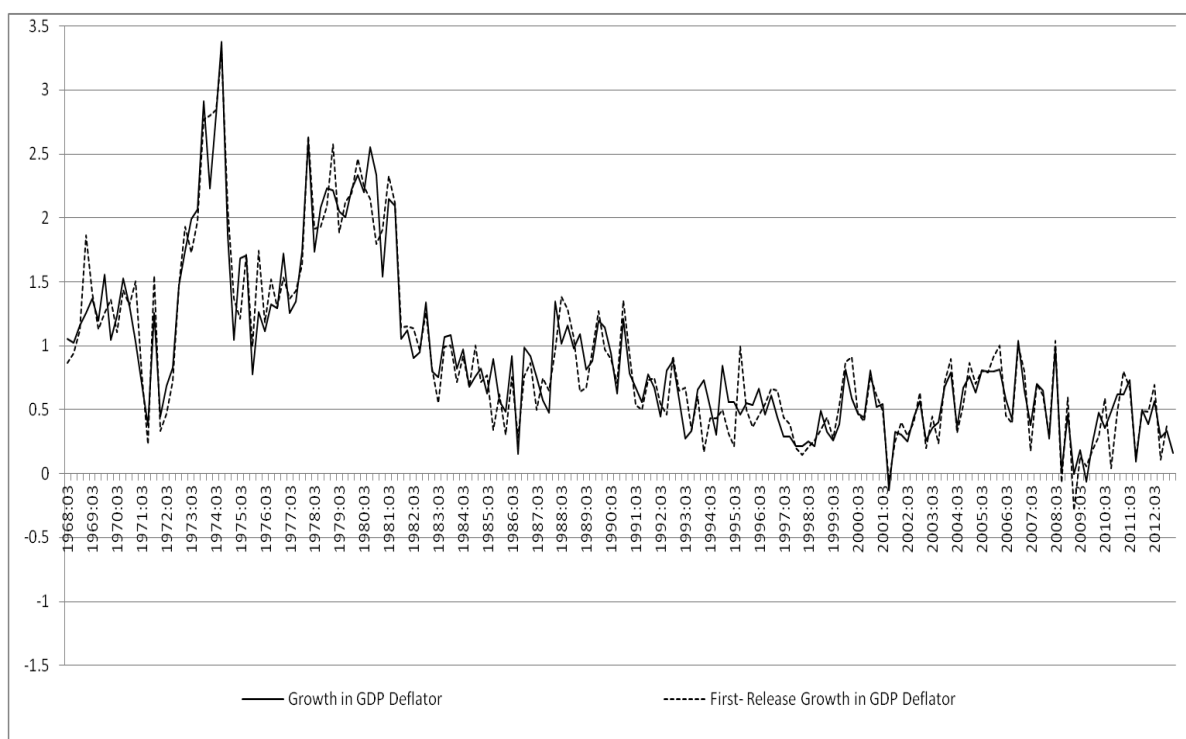


Figure A7: Actual Growth and First Release Growth in GDP Deflator defined as  ${}_T p_t - {}_T p_{t-1}$  and  ${}_{t+1} p_t - {}_t p_{t-1}$  respectively.

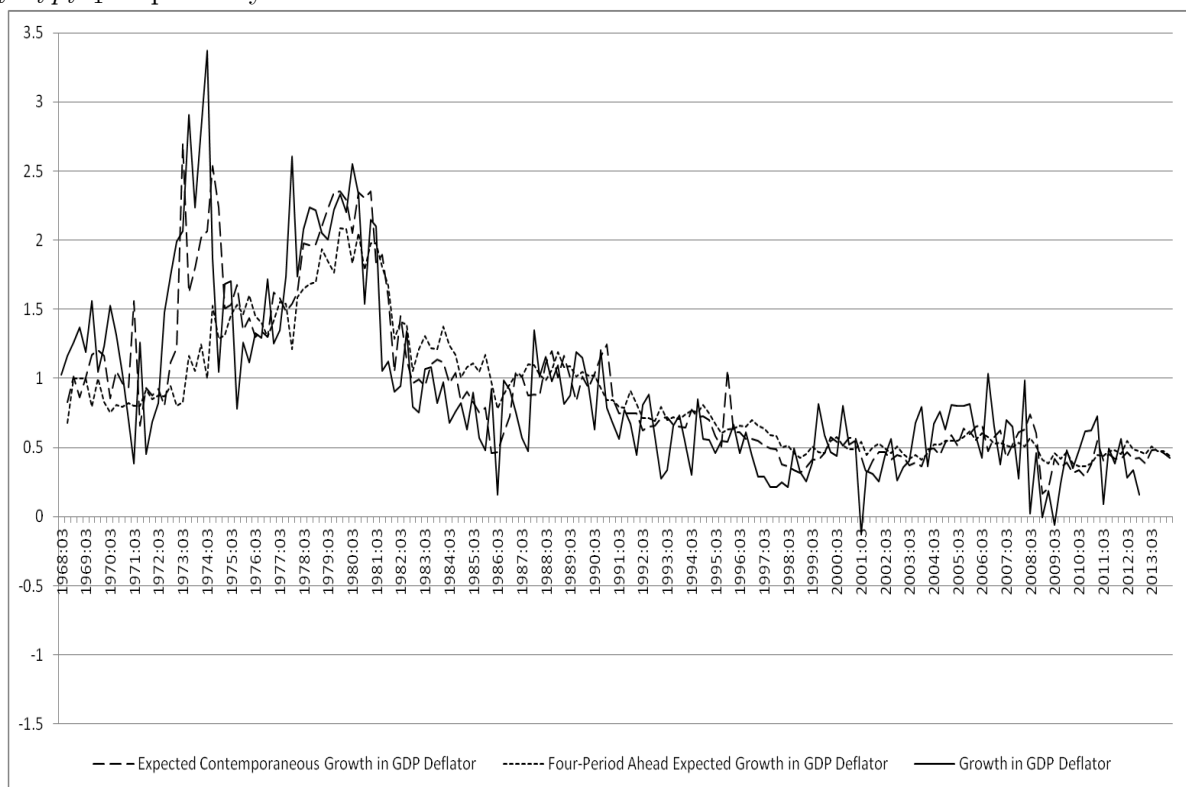


Figure A8: Expected Contemporaneous, Expected Four-Period Ahead and Actual Growth in GDP Deflator defined as  ${}_t p_t - {}_t p_{t-1}$ ,  ${}_{t-4} p_t - {}_{t-4} p_{t-1}$  and  ${}_T p_t - {}_T p_{t-1}$  respectively.

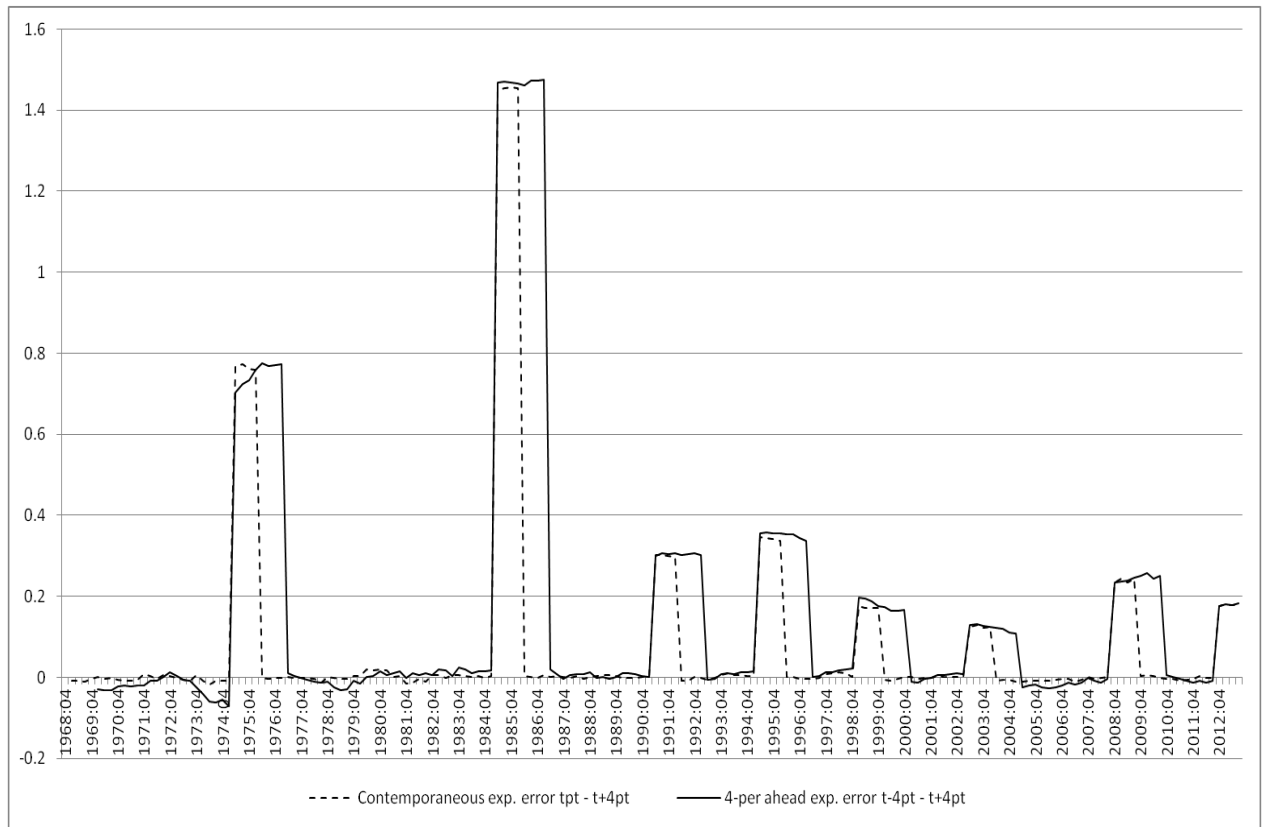


Figure A9: Contemporaneous and Four-Period Ahead Expectational Errors for the GDP Deflator Series

## 5.7 References

1. Abhyankar A., L. Sarno L. and G. Valente (2005), "Exchange rates and fundamentals: evidence on the economic value of predictability," *Journal of International Economics*, 66, pp.325-348.
2. Amisano, G. and R. Giacomini (2007), "Comparing density forecasts via weighted likelihood ratio tests," *Journal of Business and Economic Statistics*, 25, pp.177-190.
3. Andrews, D.W.K. (1993), "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, pp.821-856.
4. Andrews, D.W.K. (2003), "Tests for parameter instability and structural change with unknown change point: A corrigendum," *Econometrica*, 71, pp.395-397.

5. Ang, A., G. Bekaert and M. Wei (2007), "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?" *Journal of Monetary Economics*, 54, pp.1163-1212.
6. Aretz, K. and D.A. Peel (2010), "Spreads versus professional forecasters as predictors of future output change," *Journal of Forecasting*, 29, pp.517-522.
7. Aristidou, C., K. Lee and K. Shields (2015), "Real-Time Data should be used in Forecasting Output Growth and Recessionary Events in the US," Working Paper.
8. Arouba, S.B. (2008), "Data Revisions are not Well Behaved," *Journal of Money, Credit and Banking*, 40, pp.319-340.
9. Arouba, S.B., F. Diebold and C. Scotti (2009), "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics*, pp.417-427.
10. Bai, J. and P. Perron (1998), "Estimating and testing linear models with multiple structural Changes," *Econometrica*, 66, pp.47-78.
11. Bańbura, M. and G. Rünstler (2011), "A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP," *International Journal of Forecasting*, 27(2), pp.333-346.
12. Boero, G., Smith, J. and K.F. Wallis (2009), "Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters," *Economic Journal*, 118, pp. 1107-27.
13. Bouwman K.E. and P.A.M. Jacobs (2011), "Forecasting with real-time macroeconomic data: The ragged-edge problem and revisions," *Journal of Macroeconomics*, 33(4), pp.784-792.
14. Clark, T.E and M.W. McCracken (2010), "Averaging forecasts from VARs with uncertain instabilities," *Journal of Applied Econometrics*, 25(1), pp.5-29.
15. Clements, M. and A.B. Galvão (2013), "Real-time Forecasting of Inflation and Output Growth with autoregressive models in the Presence of Data Revisions," *Journal of Applied Econometrics*, 28(3), pp.458-477.

16. Clements, M. and D. Hendry (2005), "Guest Editors' Introduction: Information in Economic Forecasting," The MIT Press: Cambridge.
17. Croushore, D. (2006), "Forecasting with Real-Time Economic Data," in G. Elliott, C.W.J Granger and A. Timmermann (eds.) Handbook of Economic Forecasting, Amsterdam: North Holland, 961-982.
18. Croushore, D. (2010), "An Evaluation of Inflation Forecasts from Surveys using Real-Time Data," The BE Journal of Macroeconomics: Contributions, 10(1), pp.1-32.
19. Croushore, D. (2011), "Frontiers of Real-Time Data Analysis," Journal of Economic Literature, 49, pp.72-100.
20. Croushore, D. and C.L. Evans (2006), "Data revisions and the identification of monetary policy shocks," Journal of Monetary Economics, 53(6), pp.1135-1160.
21. Croushore, D. and T. Stark (2003), "A real-time data set for macroeconomists: Does the data vintage matter?" The Review of Economics and Statistics, 85(3), pp.605-617.
22. D'Agostino, A. and B. Schnatz (2012), "Survey-based nowcasting of US growth: A real-time forecast comparison over more than 40 years," European Central Bank Working Paper Series, No.1455, pp.1-22.
23. Elliott, G. and J. H. Stock (1994), "Inference in time series regression when the order of integration of a regressor is unknown," Econometric Theory, 10, pp.672-700.
24. Enders, W. (2004), "Applied Econometric Time Series," Second Edition, John Wiley & Sons: United States.
25. Faust, J. and J.H. Wright (2009), "Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset," Journal of Business and Economic Statistics, 27 (4), pp.468-479.
26. Garratt, A. and K. Lee (2010), "Investing under model uncertainty: decision based evaluation of exchange rate forecasts in the US, UK and Japan," Journal of International Money and Finance, 29(3), pp.403-422.



27. Garratt, A., K. Lee, E. Mise and K. Shields (2008), "Real time representations of the output gap," *Review of Economics and Statistics*, 90(4), pp.792-804.
28. Garratt, A., K. Lee and K. Shields (2013), "Economic Sentiment, International Interdependence and Output Dynamics in the G7," mimeo, University of Nottingham.
29. Ghysels, E. and J.H. Wright (2009), "Forecasting Professional Forecasters," *Journal of Business and Economic Statistics*, 27(4), pp.504-16.
30. Giacomini, R. and B. Rossi (2010), "Forecast comparisons in unstable environments," *Journal of Applied Econometrics*, 25(4), pp.595-620.
31. Giacomini, R. and H. White (2006), "Tests of conditional predictive ability," *Econometrica*, 74, pp.1545-1578.
32. Giannone, D., L.D., Reichlin and D. Small (2008), "Nowcasting: The Real Time Informational Content of Macroeconomic Data Releases," *Journal of Monetary Economics*, 55(4), pp.665-676.
33. Granger, C.W.J and M. Machina (2006), "Forecasting and Decision Theory," in G. Elliott, C.W.J. Granger and A.Timmermann (eds.), *Handbook of Economic Forecasting*, pp.81-98.
34. Granger, C.W.J. and M.H. Pesaran (2000), "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, 19, pp.537-560.
35. Hall, S.G. and J. Mitchell (2007), "Combining density forecasts," *Journal of Forecasting*, 23, pp. 1-13.
36. Harris, D., D.I. Harvey, S.J. Leybourne and A.M.R. Taylor (2009), "Testing for a unit root in the presence of a possible break in trend," *Econometric Theory*, 25, pp.1545-1588.
37. Harvey, D.I. and S.J. Leybourne (2013), "Break Date Estimation for Models with Deterministic Structural Change," *Oxford Bulletin of Economics and Statistics*, 0305-9049, pp.1-20.
38. Harvey, D. I., S. J Leybourne and A.M.R. Taylor (2009), "Simple, robust and powerful tests of the breaking trend hypothesis," *Econometric Theory*, Vol.25, pp.995-1029.

39. Harvey, D. I., S. J. Leybourne and A.M.R. Taylor (2010), "Robust methods for detecting multiple level breaks in autocorrelated times series," *Journal of Econometrics*, Vol.157, pp.342-358.
40. Harvey, D.I. and P. Newbold (2005), "Forecast Encompassing and Parameter Estimation," *Oxford Bulletin of Economics and Statistics* (Supplement), 67, pp.815-835.
41. Hendry, D.F. and M.P. Clements (2002), "Pooling of Forecasts," *Econometrics Journal*, 5, pp.1-26.
42. Jacobs, J. and S. van Norden (2011), "Modeling Data Revisions: Measurement Error and Dynamics of True Values," *Journal of Econometrics*, 161, 2, pp.101-109.
43. Kishor, N.K. and E.F. Koenig (2005), "VAR Estimation and Forecasting when Data are Subject to Revision," *Journal of Business and Economic Statistics*, 30(2), pp.181-190.
44. Koenig, E. F., S. Dolmas and J. Piger (2003) "The use and abuse of real-time data in economic forecasting," *The Review of Economics and Statistics*, 85(3), pp.618-628.
45. Koop, G. and S. Potter (2007), "Estimation and forecasting in models with multiple breaks," *Review of Economic Studies*, pp. 763-789.
46. Leitch, G. and J. E. Tanner (1991), "Economic Forecasts Evaluation: Profits Versus the Conventional Measures," *American Economic Review*, 81, pp.580-590.
47. Loungani, P, H. Steklerb, and N. Tamirisa (2013), "Information rigidity in growth forecasts: Some cross-country evidence", *International Journal of Forecasting*, 29, 4, pp.605-621.
48. Mankiw, G.N. and M.D. Shapiro (1986), "News or Noise? An Analysis of GNP Revisions," NBER Working Papers, 1939, National Bureau of Economic Research, Inc.
49. Matheson, T.D., J. Mitchell and B. Silverstone (2010), "Nowcasting and predicting data revisions using panel survey data," *Journal of Forecasting*, 29(3), pp.313-330.
50. Mitchell, J. and Wallis K.F. (2011), "Evaluating density forecasts: forecast combinations,

- model mixtures, calibration and sharpness," *Journal of Applied Econometrics*, 26(6), pp.1023-1040.
51. Ng, S. and P. Perron (1995), "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of American Statistical Association*, Vol. 90, Issue 429, pp.268-281.
  52. Patterson, K.D. (2002), "The Data Measurement Process for UK GNP: Stochastic Trends, Long Memory and Unit Roots," *Journal of Forecasting*, 21, pp.245-264.
  53. Perron, P. (1989), "The great crash, the oil price shock and the unit root hypothesis," *Econometrica* 57, pp.1361-1401.
  54. Perron, P. and T. Yabu (2009), "Testing for shifts in trend with an integrated or stationary noise component," *Journal of Business and Economic Statistics*, Vol.27, pp.369-396.
  55. Pesaran, M.H. and A. Timmermann (1992), "A simple nonparametric test of predictive performance," *Journal of Business and Economic Statistics*, American Association, 10(4), pp.561-65.
  56. Pesaran, M.H. and A. Timmermann (1995), "Predictability of stock returns: Robustness and Economic significance," *Journal of Finance*, American Finance Association, 50(4), pp.1201-28.
  57. Pesaran, M.H. and A. Timmermann (2007), "Selection of estimation window in the presence of breaks," *Journal of Econometrics*, 137, pp.134-161.
  58. Pesaran, M.H. and A. Timmermann (2009), "Testing dependence among serially correlated multicategory variables," *Journal of the American Statistical Association*, 104(485), pp.325-337.
  59. Pesaran, M.H. and S. Skouras (2002), "Decision-based Methods for Forecast Evaluation", in Clements, M.P. and Hendry, D.F. (eds.), *A Companion to Economic Forecasting*, Oxford: Blackwell.
  60. Pesaran, M.H., D. Pettenuzzo and A. Timmermann (2006), "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, Wiley Blackwell, vol. 73(4), pp.1057-1084.

61. Pesaran, M.H., A. Pick and M. Pranovich (2011), "Optimal Forecasts in the Presence of Structural Breaks," Nederlandsche Bank Working Paper No. 327. Available at SSRN:<http://ssrn.com/abstract=1977191> or <http://dx.doi.org/10.2139/ssrn.1977191>
  
62. Stock, J. and M.W. Watson (2006), "Forecasting with Many Predictors," in G. Elliott, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Amsterdam: North Holland, pp.515-554.
  
63. Timmermann, A. (2006), "Forecast Combinations," In Elliott, G., C.W.J. Granger and A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp. 135-196). Amsterdam: North Holland.
  
64. Vogelsang, T.J. (1998), "Testing for a shift in mean without having to estimate serial-correlation parameters", *Journal of Business and Economic Statistics*, Vol.16, pp.73-80.